The Estimation of Multi-dimensional Fixed Effects Panel Data Models

by

Laszlo Balazsi

submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of Master of Arts in Economics

Advisor: Laszlo Matyas

Budapest, Hungary

2013

Preface

Parts of chapter 2-5 have been published in the *Matyas and Balazsi*[2011/12] CEU working paper, presented at the 18*th* Panel Data Conference in Paris, and are the joint work with Laszlo Matyas. Chapter 1 and 6 are solely my own work.

Abstract

The aim of this thesis is to contribute to panel data econometrics by dealing with the most frequently used three- and four-dimensional fixed effects panel data model specifications. Within estimators are presented together with the optimal Within transformations. Also, certain data problems are taken into account, namely no self-flow and unbalanced data. Possible Within transformations are proposed here as well, to threat such data problems. Dynamic autoregressive models are also dealt with. Finally, some modifications are made in the covariance structure of the disturbances, and its implications are being investigated. The thesis ends with a brief conclusion.

Acknowledgment

First and foremost I would like to thank Laszlo Matyas for his endless efforts and his help that guided me through this thesis. Also, useful comments from the $18th$ Panel Data Conference, Paris are acknowledged. Thank my parents and Daria Pus for their continuous support.

Contents

Chapter 1

Introduction

Econometric models have always been a workhorse of empirical justification in many relevant areas in economics. Meanwhile, in the last two decades the world experienced a rapid growth in computer technology which highly affected economics itself. As the barriers of regressing models with extreme sample sizes were keep disappearing, several new possibilities presented themselves for econometricians. The reduction of computational costs had been carried over to the never ending demand for new econometric models and tools. Creating and testing such models benefits empirical researchers as well as policy makers, opening up the way for a more complex and successful analysis.

Panel data is probably the most widely applicable data structure, allowing to measure both the within group variation (between individuals) and the between group variation (between time periods). Areas where both variations carry importance are typically analyzed using panel data. The most obvious candidate is international trade (some activity between county i and j at time t) but we can also find nice examples in monetary economics.

In the last decade on panel data left its traditional 2-way (it) setup to give way to more heterogeneous ones (ijt for example). Earlier contribution to this topic can be found in *Matyas*[1997]. This new setup allows researchers for a more complex and colorful analysis but also carries the dangers and pitfalls of a not properly specified model. There are two common problems that can emerge; firstly, the number of fixed effects model specifications are growing exponentially along with the dimensions, and secondly, higher dimensional datasets are tend to be "more" incomplete. This incomplete nature of the data can distort the regression outcome if not properly addressed. It is now clear that an extension from two to three dimensions or further ahead should be treated with special attention. The main aim of this thesis is to propose new mechanisms and econometric tools to deal with such new models and the possible data problems. The thesis takes the proposed models as given, and contributes to panel data econometrics by computing the Within transformations of such fixed effects specifications together with the transformations when data problems are present, or other extensions were proposed.

In chapter 2 the most frequently used three-way fixed effects models are presented along with their Within transformations, mostly based on the work of *Egger and Pfaffermayr*[2003], *Baltagi, Egger, and Pfaffermayr*[2003], *Baldwin and Taglioni*[2006] and *Baier and Bergstrand*[2007]. It is important to note that all these models are well-used trade models carrying empirical importance. The reason why the Within transformations have extra importance is that LSDV would involve the estimation of that many parameters that the regression could even be unfeasible. By properly transforming the models the fixed effects can be eliminated as they are rarely lies in the center of the analysis. It is also shown that the Within transformations are usually not unique therefore one might choose the preferred one.

Chapter 3 investigates how these transformations should be modified if certain data problems are present, namely the lack of self-trade and unbalanced data. We speak about no self-flow if trade activity is unmeasured within a country. This means that some observations are missing from the dataset (though not randomly). After a short analysis it is easy to see that even these small gaps can result biased and inconsistent estimators. Dealing with unbalanced data is a reasonable step as several databases fell to this category: usually the individuals are not measured throughout some general time period. The results, apart from the basic algebra are supported by the work of *Wansbeek and Kapteyn*[1989]. The authors presented a solution to the traditional two-way panel data model where both individual- and time-wise fixed effects are present. Their solution is unique in a sense that the traditional order of indexing had been changed to be able to address any kind of incompletion in the data. The resulting projection matrix which removes the fixed effects can be found both in matrix and scalar form. The advantage of the latter is that inverting matrices with extreme sizes is avoided, reducing the calculations to a well-manageable amount. A great success of the *Wansbeek and Kapteyn*[1989] projection matrix is that further generalizations are possible and quite straightforward. In fact, not the dimensions of the data that matters, but the number of fixed effects included in the model: it uniquely determines the number of "iterating steps" has to be taken. With these tools in hand, unbiased estimation is now possible both in finite sample and asymptotically when the data is incomplete.

Chapter 4 examines what happens when the observed models are dynamic, namely have some sort of memory. Attention is restricted to pure dynamic autoregressive models with fixed effects, but the inclusion of other explanatory variables does not change the overall results. This chapter is mainly based on the famous results of *Nickell*[1981]. He showed that in a two-way panel data model with individual-wise fixed effects the nonzero correlation between the transformed lagged dependent variable and the transformed disturbance results biased and inconsistent estimators. This phenomenon is present in many 3-way models as well, making their estimation more compelling. Unfortunately, *Nickell*[1981] only shows the expected sign and the size of the bias, but fails to report a solution to consistently estimate the problematic models.

However the good news is that several IV and GMM estimation methods are in hand and can be widely used. One of the most popular estimator generalized by the thesis was suggested by *Arellano and Bond*[1991]. Their approach is preferable both because of its efficiency and because of the large number of orthogonality conditions used. The resulting linear GMM estimator is rather simple and also takes into account the non-scalar nature of the covariance matrix of the disturbances. Moreover, as the required orthogonality conditions easily can be found in higher dimensions, creating such estimators in a 3-dimensional setup is not that difficult. In fact, the problematic models can either be estimated directly by the *Arellano and Bond*[1991] GMM estimator or can easily be transformed in such a way that the *Arellano and Bond*[1991] estimator is then feasible.

Other generalizations can be introduced as well. Chapter 5 is designed to deal with a modified correlation structure for the disturbances: the inclusion of cross-correlation. The idea behind this is that at a given point in time some activity from country i to j and from country i to k are not necessarily independent from each other. This is a very reasonable generalization to make, resulting compelling algebra to estimate the additional correlation coefficients. As will be shown, the multiplicity of the Within transformations plays a key role here, as it provides the extra amount of identifying conditions needed to express the desired coefficients. With the estimates in hand, FGLS estimation of the models is then possible. An other finding is that some of the models fails to stand for such generalizations and the only way to deal with them is to impose some restrictions to this new setup.

There is no point in stopping at the 3-dimensional setup; it is quite easy to go further along the line and propose four-way fixed effects panel data models. Chapter 6 introduces the most frequently applied 4-dimensional models along with the optimal Within transformations. It is important to note that fixed effects in higher dimensional setups tend to depend in less indexes, giving more degrees of freedom in coming up with other transformations. One can immediately recognize its power: now certain data related problems may well be avoided only by using a different proper transformation (actually by projecting into a different subset of the four dimensional space). Just as before, the same type of data problems are being analyzed. The following conclusion can be drawn: data problems are less and less present and easy to be fixed in higher-dimensional specifications, and the same tools (*Wansbeek and Kapteyn*[1989] for example) can very well be used and generalized further.

The thesis ends with a brief conclusion. Three- and four-way fixed effects panel data specifications were being investigated. It was shown that it is rather easy to come up with the Within estimators, and also that consistent estimation is possible even if certain data problems are present. The models are flexible to other specifications (dynamic models) and to different correlation structures (inclusion of cross-correlation) as well. As the proposed estimators can very well be applied in empirical work, future usage of this thesis and the underlying working paper is possible and recommended.

Chapter 2

Models with Different Types of Heterogeneity and the Within Transformation

In three-dimensional panel data sets the dependent variable of a model is observed along three indices such as y_{ijt}, $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, and $t = 1, \ldots, T$. As in economic flows such as trade, capital (FDI), etc., there is some kind of reciprocity, we assume to start with, that $N_1 = N_2 = N$. Implicitly we also assume that the set of individuals in the observation sets i and j are the same, then we relax this assumption later on. The main question is how to formalize the individual and time heterogeneity, in our case the fixed effects. Different forms of heterogeneity yield naturally different models. In theory any fixed effects three-dimensional panel data model can directly be estimated, say for example, by least squares (LS). This involves the explicit incorporation in the model of the fixed effects through dummy variables (see for example formulation (2.16) later on). The resulting estimator is usually called Least Squares Dummy Variable (LSDV) estimator. However, it is well known that the first moment of the LS estimators is invariant to linear transformations, as long as the transformed explanatory variables and disturbance terms remain uncorrelated. So if we could transform the model, that is all variables of the model, in such a way that the transformation wipes out the fixed effects, and then estimate this transformed model by least squares, we would get parameter estimates with similar first moment properties (unbiasedness) as those from the estimation of the original untransformed model. This would be simpler as the fixed effects then need not to be estimated or explicitly incorporated into the model[2]. We must emphasize, however, that these transformations are usually not unique in our context. The resulting different Within estimators (for the same model), although have the same bias/unbiasedness, may not give numerically the same parameter estimates. This comes from the fact that the different Within transformations represent different projection in the (i, j, t) space, so the corresponding Within estimators may in fact use different subsets of the three-dimensional data space. Due to the Gauss-Markov and the Frisch-Waugh theorems (see, for example, *Gourieroux and Monfort*[1989]), there is always an optimal Within estimator, exactly the one which is based on the transformations generated by the appropriate LSDV estimator.

[2]An early partial overview of these transformations can be found in *Laszlo Matyas and Konya*[2011/05].

Why to bother then, and not always use the LSDV estimator directly? First, because when the data becomes larger, the estimation of a model with the fixed effects explicitly incorporated into it is quite difficult, or even practically impossible, so the use of Within estimators can be quite useful. Then, we may also exploit the different projections and the resulting various Within estimators to deal with some data generated problems.

The first attempt the properly extend the standard fixed effects panel data model (see for example *Baltagi*[1995] or *Balestra and Krishnakumar*[2008]) to a multidimensional setup was proposed by *Matyas*[1997]. The specification of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt} \qquad i = 1, \ldots, N \quad j = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2.1}$$

where the α, γ and λ parameters are time and country specific fixed effects, the x variables are the usual covariates, β $(K \times 1)$ the focus structural parameters and ε is the idiosyncratic disturbance term, for which (unless otherwise stated)

$$E(\varepsilon_{ijt}) = 0, \qquad E(\varepsilon_{ijt}\varepsilon_{i'j't'}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } i = i',\ j = j' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

and we also assume that the covariates and the disturbance terms are uncorrelated.

The simplest Within transformation for this model is

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_t + \bar{y}) \tag{2.3}$$

where

$$\bar{y}_{ij} = 1/T \sum_t y_{ijt}$$
$$\bar{y}_t = 1/N^2 \sum_i \sum_j y_{ijt}$$
$$\bar{y} = 1/N^2 T \sum_i \sum_j \sum_t y_{ijt}$$

However, the optimal Within transformation (which actually gives numerically the same parameter estimates as the direct LS estimation of model (2.1), that is the LSDV estimator) is in fact

$$(y_{ijt} - \bar{y}_i - \bar{y}_j - \bar{y}_t + 2\bar{y}) \tag{2.4}$$

where

$$\bar{y}_i = 1/(NT) \sum_j \sum_t y_{ijt}$$
$$\bar{y}_j = 1/(NT) \sum_i \sum_t y_{ijt}$$

Let us note here that this model is suited to deal with purely cross sectional data as well (that is when $T = 1$). In this case, there are only the α_i and γ_j fixed effects and the the appropriate Within transformation is $(y_{ij} - \bar{y}_j - \bar{y}_i + \bar{y})$ with $\bar{y} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_{ij}$.

Another model has been proposed by *Egger and Pfaffermayr*[2003] which takes into account bilateral interaction effects. The model specification is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \varepsilon_{ijt} \tag{2.5}$$

where the γ_{ij} are the bilateral specific fixed effects (this approach can easily be extended to account for multilateral effects as well). The simplest (and optimal) Within transformation which clears the fixed effects now is

$$(y_{ijt} - \bar{y}_{ij}) \quad \text{where} \quad \bar{y}_{ij} = 1/T \sum_t y_{ijt} \tag{2.6}$$

It can be seen that the use of the Within estimator here, and even more so for the models discussed later, is highly recommended as direct estimation of the model by LS would involve the estimation of $(N \times N)$ parameters which is not very practical for larger N. For model (2.14) this would even be practically impossible.

A variant of model (2.5) often used in empirical studies is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \lambda_t + \varepsilon_{ijt} \tag{2.7}$$

As model (2.1) is in fact a special case of this model (2.7), transformation (2.3) can be used to clear the fixed effects. While transformation (2.3) leads to the optimal Within estimator for model (2.7), it is clear why it is not optimal for model (2.1): it "over-clears" the fixed effects, as it does not take into account the parameter restrictions $\gamma_{ij} = \alpha_i + \gamma_i$. It is worth noticing that models (2.5) and (2.7) are in fact straight panel data models where the individuals are now the (ij) pairs.

Several other forms of fixed effects were suggested by *Baltagi, Egger, and Pfaffermayr*[2003], *Baldwin and Taglioni*[2006] and *Baier and Bergstrand*[2007]. A simpler model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{jt} + \varepsilon_{ijt} \tag{2.8}$$

The Within transformation which clears the fixed effects is

$$(y_{ijt} - \bar{y}_{jt}) \quad \text{where} \quad \bar{y}_{jt} = 1/N \sum_i y_{ijt} \tag{2.9}$$

Another variant of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \varepsilon_{ijt} \tag{2.10}$$

Here the Within transformation which clears the fixed effects is

$$(y_{ijt} - \bar{y}_{it}) \quad \text{where} \quad \bar{y}_{it} = 1/N \sum_j y_{ijt} \tag{2.11}$$

The most frequently used variation of this model is

$$y_{ijt} = \beta' x_{ijt} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt} \tag{2.12}$$

The required Within transformation here is

$$(y_{ijt} - 1/N \sum_i y_{ijt} - 1/N \sum_j y_{ijt} + 1/N^2 \sum_i \sum_j y_{ijt})$$

or in short

$$(y_{ijt} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t) \tag{2.13}$$

Let us notice here that transformation (2.13) clears the fixed effects for model (2.1) as well, but of course the resulting Within estimator is not optimal. The model which encompasses all above effects is

$$y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt} \tag{2.14}$$

By applying suitable restrictions to model (2.14) we can obtain the models discussed above. The Within transformation for this model is

$$\begin{aligned}(y_{ijt} \quad & -1/T \textstyle\sum_t y_{ijt} - 1/N \sum_i y_{ijt} - 1/N \sum_j y_{ijt} + 1/N^2 \sum_i \sum_j y_{ijt} \\ & +1/(NT) \textstyle\sum_i \sum_t y_{ijt} + 1/(NT) \sum_j \sum_t y_{ijt} - 1/(N^2T) \sum_i \sum_j \sum_t y_{ijt})\end{aligned}$$

or in a shorter form

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y}) \tag{2.15}$$

We can write up these Within transformations in a more compact matrix form using *Davis*[2002]'s and *Hornok*[2011]'s approach. Model (2.14) in matrix form is

$$y = X\beta + \tilde{D}_1\gamma + \tilde{D}_2\alpha + \tilde{D}_3\alpha_* + \varepsilon \tag{2.16}$$

where $y, (N^2T \times 1)$ is the vector of the dependent variable, $X, (N^2T \times K)$ is the matrix of explanatory variables, γ, α and α_* are the vectors of fixed effects with size $(N^2T \times N^2)$, $(N^2T \times NT)$ and $(N^2T \times NT)$ respectively,

$$\tilde{D}_1 = I_{N^2} \otimes l_T, \quad \tilde{D}_2 = I_N \otimes l_N \otimes I_T \quad \text{and} \quad \tilde{D}_3 = l_N \otimes I_{NT}$$

l is the vector of ones and I is the identity matrix with the appropriate size in the index. Let $D = (\tilde{D}_1, \tilde{D}_2, \tilde{D}_3)$, $Q_D = D(D'D)^{-1}D'$ and $P_D = I - Q_D$. Using *Davis*[2002]'s method it can be shown that $P_D = P_1 - Q_2 - Q_3$ where

$$\begin{aligned}
P_1 &= (I_N - \bar{J}_N) \otimes I_{NT} \\
Q_2 &= (I_N - \bar{J}_N) \otimes \bar{J}_N \otimes I_T \\
Q_3 &= (I_N - \bar{J}_N) \otimes (I_N - \bar{J}_N) \otimes \bar{J}_T \\
\bar{J}_N &= \tfrac{1}{N} J_N, \quad \bar{J}_T = \tfrac{1}{T} J_T
\end{aligned}$$

and J is the matrix of ones with its size in the index. Collecting all these terms we get

$$\begin{aligned}
P_D &= \left[(I_N - \bar{J}_N) \otimes (I_N - \bar{J}_N) \otimes (I_T - \bar{J}_T)\right] \\
&= I_{N^2T} - (\bar{J}_N \otimes I_{NT}) - (I_N \otimes \bar{J}_N \otimes I_T) - (I_{N^2} \otimes \bar{J}_T) \\
&+ (I_N \otimes \bar{J}_{NT}) + (\bar{J}_N \otimes I_N \otimes \bar{J}_T) + (\bar{J}_{N^2} \otimes I_T) - \bar{J}_{N^2T}
\end{aligned}$$

The typical element of P_D gives the transformation (2.15). By appropriate restrictions on the parameters of (2.16) we get back the previously analysed Within transformations.

Now transforming model (2.16) with transformation (2.15) leads to

$$\underbrace{P_D y}_{y_p} = \underbrace{P_D X}_{X_p} \beta + \underbrace{P_D \tilde{D}_1}_{=0} \gamma + \underbrace{P_D \tilde{D}_2}_{=0} \alpha + \underbrace{P_D \tilde{D}_3}_{=0} \alpha_* + \underbrace{P_D \varepsilon}_{\varepsilon_p}$$

and the corresponding Within estimator is

$$\widehat{\beta}_W = (X_p' X_p)^{-1} X_p y_p$$

This in fact is the optimal estimator as P_D is the Frisch-Waugh projection matrix, implying the optimality of $\widehat{\beta}_W$.

Chapter 3

Some Data Problems

3.1 No Self Flow Data

As these multidimensional panel data models are frequently used to deal with flow types of data like trade, capital movements (FDI), etc., it is important to have a closer look at the case when, by nature, we do not observe self flow. This means that from the (ijt) indexes we do not have observations for the dependent variable of the model when $i = j$ for any t. This is the first step to relax our initial assumption that $N_1 = N_2 = N$ and that the observation sets i and j are equivalent.

For most of the previously introduced models this is not a problem, the Within transformations work as they are meant to and eliminate the fixed effects. However, this is not the case unfortunately for models (2.1) (transformation (2.4)), (2.12) and (2.14). Let us have a closer look at the difficulty. For model (2.1) and transformation (2.4), instead of canceled out fixed effects, we end up with the following remaining fixed effects

$$\begin{aligned}\alpha_i^* &= \alpha_i - \tfrac{1}{(N-1)T}\cdot(N-1)T\cdot\alpha_i - \tfrac{1}{(N-1)T}\textstyle\sum_{i=1;\, i\neq j}^{N} T\cdot\alpha_i \\ &\quad -\tfrac{1}{N(N-1)}\textstyle\sum_{i=1}^{N}(N-1)\cdot\alpha_i + \tfrac{2}{N(N-1)T}\textstyle\sum_{i=1}^{N}(N-1)T\cdot\alpha_i \\ &= \alpha_i - \alpha_i - \tfrac{1}{N-1}\textstyle\sum_{i=1;\, i\neq j}^{N}\alpha_i + \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\alpha_i = \tfrac{1}{N}\alpha_j - \tfrac{1}{N(N-1)}\textstyle\sum_{i=1;\, i\neq j}^{N}\alpha_i\end{aligned}$$

$$\begin{aligned}\gamma_j^* &= \gamma_j - \tfrac{1}{(N-1)T}\textstyle\sum_{j=1;\, j\neq i}^{N} T\cdot\gamma_j - \tfrac{1}{(N-1)T}\cdot(N-1)T\cdot\gamma_j \\ &\quad -\tfrac{1}{N(N-1)}\textstyle\sum_{j=1}^{N}(N-1)\cdot\gamma_j + \tfrac{2}{N(N-1)T}\textstyle\sum_{j=1}^{N}(N-1)T\cdot\gamma_j \\ &= \gamma_j - \tfrac{1}{N-1}\textstyle\sum_{j=1;\, j\neq i}^{N}\gamma_j - \gamma_j + \tfrac{1}{N}\textstyle\sum_{j=1}^{N}\gamma_j = \tfrac{1}{N}\gamma_i - \tfrac{1}{N(N-1)}\textstyle\sum_{j=1;\, j\neq i}^{N}\gamma_j\end{aligned}$$

and for the time effects

$$\begin{aligned}\lambda_t^* &= \lambda_t - \tfrac{1}{(N-1)T}\textstyle\sum_{t=1}^{T}(N-1)\cdot\lambda_t - \tfrac{1}{(N-1)T}\textstyle\sum_{t=1}^{T}(N-1)\cdot\lambda_t \\ &\quad -\tfrac{1}{N(N-1)}\cdot N(N-1)\lambda_t + \tfrac{2}{N(N-1)T}\textstyle\sum_{t=1}^{T}N(N-1)\cdot\lambda_t = \\ &= \lambda_t - \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\lambda_t - \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\lambda_t - \lambda_t + \tfrac{2}{T}\textstyle\sum_{t=1}^{T}\lambda_t = 0\end{aligned}$$

So clearly this Within estimator now is biased. The bias is of course eliminated if we add the (ii) observations back to the above bias formulae, and also, quite intuitively, when $N \to \infty$. On the other hand, luckily, transformation (2.3) as seen earlier, although not optimal, leads to an unbiased Within estimator for model (2.1) and remains so even in the lack of self flow data.

As seen earlier model (2.1) is suited to deal with the purely cross sectional case. Then, however, the appropriate Within transformation that clears the fixed effects is in fact

$$y_{ij} - \bar{y}_j - \bar{y}_i + \frac{N}{N-1}\bar{y} - \frac{1}{N-1}y_{ji}$$

Now, let us continue with model (2.12). After the Within transformation (2.13), instead of canceled out fixed effects we end up with the following remaining fixed effects

$$\begin{aligned}\alpha^*_{it} &= \alpha_{it} - \tfrac{1}{N-1}\textstyle\sum_{i=1;i\neq j}^{N}\alpha_{it} - \tfrac{1}{N-1}(N-1)\alpha_{it} + \tfrac{1}{N(N-1)}\sum_{i=1}^{N}(N-1)\alpha_{it} \\ &= -\tfrac{1}{N(N-1)}\textstyle\sum_{k=1;k\neq j}^{N}\alpha_{kt} + \tfrac{1}{N}\alpha_{jt}\end{aligned}$$

and

$$\begin{aligned}\gamma^*_{jt} &= \gamma_{jt} - \tfrac{1}{N-1}(N-1)\gamma_{jt} - \tfrac{1}{N-1}\textstyle\sum_{j=1;j\neq i}^{N}\gamma_{jt} + \tfrac{1}{N(N-1)}\sum_{j=1}^{N}(N-1)\gamma_{jt} \\ &= -\tfrac{1}{N(N-1)}\textstyle\sum_{l=1;l\neq i}^{N}\gamma_{lt} + \tfrac{1}{N}\gamma_{it}\end{aligned}$$

As long as the α^* and γ^* parameters are not zero, the Within estimators will be biased. Similarly for model (2.14), the remaining fixed effects are now

$$\begin{aligned}\gamma^*_{ij} &= \gamma_{ij} - \tfrac{1}{T}T\cdot\gamma_{ij} - \tfrac{1}{N-1}\textstyle\sum_{i=1;i\neq j}^{N}\gamma_{ij} - \tfrac{1}{N-1}\sum_{j=1;j\neq i}^{N}\gamma_{ij} \\ &+\tfrac{1}{N(N-1)}\textstyle\sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}\gamma_{ij} + \tfrac{1}{(N-1)T}\sum_{i=1;i\neq j}^{N}T\gamma_{ij} \\ &+\tfrac{1}{(N-1)T}\textstyle\sum_{j=1;j\neq i}^{N}T\gamma_{ij} - \tfrac{1}{N(N-1)T}\sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}T\gamma_{ij} = 0\end{aligned}$$

but

$$\begin{aligned}
\alpha^*_{it} &= \alpha_{it} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\alpha_{it} - \tfrac{1}{N-1}\sum_{i=1;i\neq j}^{N}\alpha_{it} - \tfrac{1}{N-1}(N-1)\alpha_{it} \\
&+\tfrac{1}{N(N-1)}\textstyle\sum_{i=1}^{N}(N-1)\alpha_{it} + \tfrac{1}{(N-1)T}\sum_{i=1;i\neq j}^{N}\sum_{t=1}^{T}\alpha_{it} \\
&+\tfrac{1}{(N-1)T}\textstyle\sum_{t=1}^{T}(N-1)\alpha_{it} - \tfrac{1}{N(N-1)T}\sum_{i=1}^{N}\sum_{t=1}^{T}(N-1)\alpha_{it} \\
&= \tfrac{1}{N(N-1)T}\textstyle\sum_{i=1;i\neq j}^{N}\sum_{t=1}^{T}\alpha_{it} + \tfrac{1}{NT}\sum_{t=1}^{T}\alpha_{jt} - \tfrac{1}{N(N-1)}\sum_{i=1;i\neq j}^{N}\alpha_{it} + \tfrac{1}{N}\alpha_{jt}
\end{aligned}$$

and, finally

$$\begin{aligned}
\tilde{\alpha}^*_{jt} &= \tilde{\alpha}_{jt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\tilde{\alpha}_{jt} - \tfrac{1}{N-1}(N-1)\tilde{\alpha}_{jt} - \tfrac{1}{N-1}\sum_{j=1;j\neq i}^{N}\tilde{\alpha}_{jt} \\
&+\tfrac{1}{N(N-1)}\textstyle\sum_{j=1}^{N}(N-1)\tilde{\alpha}_{jt} + \tfrac{1}{(N-1)T}\sum_{t=1}^{T}(N-1)\tilde{\alpha}_{jt} \\
&+\tfrac{1}{(N-1)T}\textstyle\sum_{j=1;j\neq i}^{N}\sum_{t=1}^{T}\tilde{\alpha}_{jt} - \tfrac{1}{N(N-1)T}\sum_{j=1}^{N}\sum_{t=1}^{T}(N-1)\tilde{\alpha}_{jt} \\
&= \tfrac{1}{N(N-1)T}\textstyle\sum_{j=1;j\neq i}^{N}\sum_{t=1}^{T}\tilde{\alpha}_{jt} + \tfrac{1}{NT}\sum_{t=1}^{T}\tilde{\alpha}_{it} - \tfrac{1}{N(N-1)}\sum_{j=1;j\neq i}^{N}\tilde{\alpha}_{jt} + \tfrac{1}{N}\tilde{\alpha}_{it}
\end{aligned}$$

where in order to avoid confusion with the two similar α fixed effects α_{jt} is now denoted by $\tilde{\alpha}_{jt}$. It can be seen, as expected, these remaining fixed effects are indeed wiped out when ii type observations are present in the data. When $N \to \infty$ the remaining effects go to zero, which implies that the bias of the Within estimators go to zero as well.

Fortunately, however, there is good news as well. For both models (2.12) and (2.14) there is a transformation which wipes out the fixed effects, and so remains unbiased even in this case. For model (2.12) this can be written up as

$$\begin{aligned} y_{ijt} & -\bar{y}_{it} - \bar{y}_{jt} + \bar{y}_t + \tfrac{1}{N-1}\bar{y}_t - \tfrac{1}{N-1}y_{jit} = \\ y_{ijt} & -\bar{y}_{it} - \bar{y}_{jt} + \tfrac{N}{N-1}\bar{y}_t - \tfrac{1}{N-1}y_{jit} \end{aligned} \tag{3.1}$$

or in matrix form

$$\begin{aligned} &\big(I_{N(N-1)T} - I_N \otimes \bar{J}_{N-1} \otimes I_T - \bar{J}_{N-1} \otimes I_{NT} + \bar{J}_{N(N-1)} \otimes I_T \\ &\quad + \tfrac{1}{N-1}\bar{J}_{N(N-1)} \otimes I_T - \tfrac{1}{N-1}K_{N(N-1)} \otimes I_T\big) = \\ &\big(I_{N(N-1)T} - I_N \otimes \bar{J}_{N-1} \otimes I_T - \bar{J}_{N-1} \otimes I_{NT} + \tfrac{N}{N-1}\bar{J}_{N(N-1)} \otimes I_T \\ &\quad - \tfrac{1}{N-1}K_{N(N-1)} \otimes I_T\big) \end{aligned}$$

where $K_{N(N-1)}$ is the matrix with the following rows: the row corresponding to observation ij is a row of 0-s with 1 in the ji th place, that is the ij th row is in fact

$$\Big[0, 0, \dots 0, \underbrace{1}_{ji-\text{th element}}, 0, \dots, 0\Big]$$

For model (2.14) the appropriate transformation is

$$\begin{aligned} y_{ijt} & -\bar{y}_{jt} - \bar{y}_{it} - \bar{y}_{ij} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y} - \tfrac{1}{N-1}\bar{y} + \tfrac{1}{N-1}\bar{y}_t + \tfrac{1}{N-1}\bar{y}_{ji} - \tfrac{1}{N-1}y_{jit} = \\ y_{ijt} & -\bar{y}_{jt} - \bar{y}_{it} - \bar{y}_{ij} + \tfrac{N}{N-1}\bar{y}_t + \bar{y}_j + \bar{y}_i - \tfrac{N}{N-1}\bar{y} + \tfrac{1}{N-1}\bar{y}_{ji} - \tfrac{1}{N-1}y_{jit} \end{aligned} \tag{3.2}$$

or, again, in matrix form

$$
\begin{aligned}
&\big(I_{N(N-1)T} - \bar{J}_{N-1} \otimes I_{NT} - I_N \otimes \bar{J}_{N-1} \otimes I_T - I_{N(N-1)} \otimes \bar{J}_T \\
&\quad + \bar{J}_{N(N-1)} \otimes I_T + \bar{J}_{N-1} \otimes I_N \otimes \bar{J}_T + I_N \otimes \bar{J}_{(N-1)T} - \bar{J}_{N(N-1)T} \\
&\quad - \tfrac{1}{N-1}\bar{J}_{N(N-1)T} + \tfrac{1}{N-1}\bar{J}_{N(N-1)} \otimes I_T + \tfrac{1}{N-1}K_{N(N-1)} \otimes \bar{J}_T \\
&\quad - \tfrac{1}{N-1}K_{N(N-1)} \otimes I_T\big) = \\
&\big(I_{N(N-1)T} - \bar{J}_{N-1} \otimes I_{NT} - I_N \otimes \bar{J}_{N-1} \otimes I_T - I_{N(N-1)} \otimes \bar{J}_T \\
&\quad + \tfrac{N}{N-1}\bar{J}_{N(N-1)} \otimes I_T + \bar{J}_{N-1} \otimes I_N \otimes \bar{J}_T + I_N \otimes \bar{J}_{(N-1)T} - \tfrac{N}{N-1}\bar{J}_{N(N-1)T} \\
&\quad + \tfrac{1}{N-1}K_{N(N-1)} \otimes \bar{J}_T - \tfrac{1}{N-1}K_{N(N-1)} \otimes I_T\big)
\end{aligned}
$$

So overall, the self flow data problem can be overcome by using an appropriate Within transformation leading to an unbiased estimator.

Next, we can go further along the above lines and see what going is to happen if the observation sets i and j are different. If the two sets are completely disjunct, say for example if we are modeling export activity between the EU and APEC countries, intuitively enough, for all the models considered the Within estimators are unbiased, even in finite samples, as the no-self-trade problem do not arise. If the two sets are not completely disjunct, on the other hand, say for example in the case of trade between the EU and OECD countries, as the no-self-trade do arise, we are face with the same biases outlined above. Unfortunately, however, transformations (3.1) and (3.2) do not work in this case, and there are no obvious transformations that could be worked out for this scenario.

3.2 Unbalanced Data

Like in the case of the usual panel data sets (see *Wansbeek and Kapteyn*[1989] or *Baltagi*[1995], for example), just more frequently, one may be faced with the situation when the data at hand is unbalanced. In our framework of analysis this means that for all the previously studied models, in general $t = 1, \ldots, T_{ij}$, $\sum_i \sum_j T_{ij} = T$ and T_{ij} is often not equal to $T_{i'j'}$. For models (2.5), (2.8), (2.10) and (2.12) the unbalanced nature of the data does not cause any problems, the Within transformations can be used, and have exactly the same properties, as in the balanced case. However, for models (2.1) and (2.14) we are facing trouble.

In the case of model (2.1) and transformation (2.3) we get for the fixed effects the following terms (let us remember: this in fact is the optimal transformation for model (2.7))

$$\begin{aligned} \alpha_i^* &= \alpha_i - \tfrac{1}{T_{ij}} \textstyle\sum_{t=1}^{T_{ij}} \alpha_i - \tfrac{1}{N^2} \sum_{i=1}^{N} N\alpha_i + \tfrac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}} \alpha_i \\ &= -\tfrac{1}{N} \textstyle\sum_{i=1}^{N} \alpha_i + \tfrac{1}{T} \sum_{i=1}^{N} \left(\alpha_i \cdot \sum_{j=1}^{N} T_{ij} \right) \\ &= \tfrac{1}{NT} \textstyle\sum_{i=1}^{N} \alpha_i \cdot (N \sum_{j=1}^{N} T_{ij} - T) \end{aligned}$$

$$\begin{aligned}\gamma_j^* &= \gamma_j - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\gamma_j - \tfrac{1}{N^2}\sum_{j=1}^{N} N\gamma_j + \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_j \\ &= -\tfrac{1}{N}\textstyle\sum_{j=1}^{N}\gamma_j + \tfrac{1}{T}\sum_{j=1}^{N}\left(\gamma_j\cdot\sum_{i=1}^{N}T_{ij}\right) \\ &= \tfrac{1}{NT}\textstyle\sum_{j=1}^{N}\gamma_j\cdot\left(N\sum_{i=1}^{N}T_{ij} - T\right)\end{aligned}$$

and

$$\begin{aligned}\lambda_t^* &= \lambda_t - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\lambda_t - \tfrac{1}{N^2}N^2\lambda_t + \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\lambda_t \\ &= \lambda_t - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\lambda_t - \lambda_t + \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\lambda_t \\ &= -\tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\lambda_t + \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\lambda_t\end{aligned}$$

These terms clearly do not add up to zero in general, so the Within transformation does not clear the fixed effects, as a result this Within estimator will be biased. (It can easily checked that the above α_i^*, γ_j^* and λ_t^* terms add up to zero when $\forall i, j\ T_{ij} = T$.) As (2.3) is the optimal Within estimator for model (2.7), this is bad news for the estimation of that model as well. We, unfortunately, get very similar results for transformation (2.4) too. The good news is, on the other hand, as seen earlier, that for model (2.1) transformation (2.13) clears the fixed effects, and although not optimal in this case, it does not depend on time, so in fact the corresponding Within estimator is still unbiased in this case.

Unfortunately, no such luck in the case of model (2.14) and transformation (2.15). The remaining fixed effects are now

$$
\begin{aligned}
\gamma_{ij}^* &= \gamma_{ij} - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\gamma_{ij} - \tfrac{1}{N}\sum_{i=1}^{N}\gamma_{ij} - \tfrac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}+\\
&\quad + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij} + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\gamma_{ij}\\
&= \gamma_{ij} - \gamma_{ij} - \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\gamma_{ij} - \tfrac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij}+\\
&\quad + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\textstyle\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}\\
&= -\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\gamma_{ij} - \tfrac{1}{N}\sum_{j=1}^{N}\gamma_{ij} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\gamma_{ij}T_{ij}+\\
&\quad + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\textstyle\sum_{j=1}^{N}\gamma_{ij}T_{ij} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}T_{ij}
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{it}^* &= \alpha_{it} - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\alpha_{it} - \tfrac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \tfrac{1}{N}\sum_{j=1}^{N}\alpha_{it} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{it}+\\
&\quad + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}\\
&= \alpha_{it} - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\alpha_{it} - \tfrac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \alpha_{it} + \tfrac{1}{N}\sum_{i=1}^{N}\alpha_{it}+\\
&\quad + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}\\
&= -\tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\alpha_{it} + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it} + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}-\\
&\quad -\tfrac{1}{T}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{it}
\end{aligned}
$$

and

$$\begin{aligned}\alpha^*_{jt} &= \alpha_{jt} - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\alpha_{jt} - \tfrac{1}{N}\sum_{i=1}^{N}\alpha_{jt} - \tfrac{1}{N}\sum_{j=1}^{N}\alpha_{jt} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{jt}+ \\ &\quad + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} \\ &= \alpha_{jt} - \tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\alpha_{jt} - \alpha_{jt} - \tfrac{1}{N}\sum_{i=1}^{N}\alpha_{jt} + \tfrac{1}{N}\sum_{i=1}^{N}\alpha_{jt} + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt}+ \\ &\quad + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\textstyle\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} - \tfrac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} \\ &= -\tfrac{1}{T_{ij}}\textstyle\sum_{t=1}^{T_{ij}}\alpha_{jt} + \tfrac{1}{\sum_{i=1}^{N}T_{ij}}\sum_{i=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt} + \tfrac{1}{\sum_{j=1}^{N}T_{ij}}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt}- \\ &\quad -\tfrac{1}{T}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_{ij}}\alpha_{jt}\end{aligned}$$

These terms clearly do not cancel out in general, as a result the corresponding Within estimator is biased. Unfortunately, the increase of N does not deal with the problem, so the bias remains even when $N \to \infty$. It can easily be checked, however, that in the balanced case, i.e., when each $T_{ij} = T/N^2$ the fixed effects drop out indeed from the above formulations. Therefore, from a practical point of view, the estimation of both models (2.7) and (2.14) is quite problematic. However, luckily, the *Wansbeek and Kapteyn*[1989] approach can be extended to these cases. In the case of model (2.7), picking up the notation used in (2.16), $\tilde{D}_1$ and $\tilde{D}_2$ have to be modified to reflect the unbalanced nature of the data. Recall that t goes from 1 to some T_{ij}, and we assume $\sum_{ij} T_{ij} \equiv T$ and let $max\{T_{ij}\} \equiv T^*$ Then let the V_t-s be the sequence of I_{N^2} matrixes, $(t = 1 \dots T^*)$ in which the following adjustments were made: for each ij observation, we leave the row (representing ij) in the first T_{ij} matrixes untouched, but delete them from the remaining

$T^* - T_{ij}$ matrixes. In this way we end up with the following dummy variable setup

$$\begin{aligned} D_1 &= [V_1', V_2' \ldots V_{T^*}']' \,, \ (T \times N^2) \,; \\ D_2^a &= diag \left\{ V_1 \cdot l_{N^2}, V_2 \cdot l_{N^2} \ldots, V_{T^*} \cdot l_{N^2} \right\} \,, \ (T \times T^*); \end{aligned}$$

So the complete dummy variable structure now is $D^a = (D_1, D_2^a)$. Let us note here, that in this case, just as in *Wansbeek and Kapteyn*[1989], index t goes "slowly" and ij "fast".

Let now

$$\Delta_{N^2} \equiv D_1' D_1 \,, \qquad \Delta_{T^*} \equiv D_2^{a'} D_2^a \,, \qquad A^a \equiv D_2^{a'} D_1^a \,,$$

and

$$\begin{aligned} \bar{D}^a &\equiv D_2^a - D_1 \Delta_{N^2}^{-1} A^{a'} = \left(I_T - D_1 \left(D_1' D_1 \right)^{-1} D_1' \right) D_2^a \\ Q^a &\equiv \Delta_{T^*} - A^a \Delta_{N^2}^{-1} A^{a'} = D_2^{a'} \bar{D}^a \end{aligned}$$

Note that in the original balanced case $\Delta_{N^2} = T \cdot I_{N^2}$, $\Delta_{T^*} = N^2 \cdot I_T$ and $A^a = l_T \otimes l_{N^2}'$. So finally, the appropriate transformation for model (2.7) is

$$P^a = \left(I_T - D_1 \Delta_{N^2}^{-1} D_1' \right) - \bar{D}^a Q^{a-} \bar{D}^{a'} \tag{3.3}$$

where Q^{a-} denotes the generalized inverse, as, like in case of *Wansbeek and Kapteyn*[1989], the Q^a matrix has no full rank. We can re-write transformation (3.3) using scalar notation for the ease of computation. For y let $\bar{\phi}^a \equiv Q^{a-} \bar{D}^{a'} y$. In that way, a particular element

(ijt) of $P^a y$ can be written up as

$$[P^a y]_{ijt} = y_{ijt} - \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} y_{ijt} - \bar{\phi}_t^a + \frac{1}{T_{ij}} a_{ij}^{a'} \bar{\phi}^a,$$

where a_{ij}^a is the ij-th column of matrix A^a (A^a has N^2 columns), and $\bar{\phi}_t^a$ is the t-th element of the $(T^* \times 1)$ column vector $\bar{\phi}^a$. (Note that we only have to calculate the inverse of a $(T^* \times T^*)$ matrix, which is easily doable.)

Let us continue with model (2.14) and let now the matrix of dummy variables for the fixed effects be $D^b = (D_1, D_2^b, D_3)$ where D_1 is defined as above,

$$D_2^b = diag\{U_1, \ldots U_{T^*}\}$$

with the U_t-s being the $I_N \otimes l_N$ matrixes at time t but modified in the following way: we leave untouched the rows corresponding to observation ij in the first T_{ij} matrix, but delete them from the other $T^* - T_{ij}$ matrixes, and

$$D_3 = diag\{W_1, \ldots W_{T^*}\}$$

with the W_t-s being the $l_N \otimes I_N$ matrixes at time t, with the same modifications as above.

Defining the partial projector matrixes B and C as

$$\begin{aligned} B &\equiv I_T - D_1(D_1'D_1)^{-1}D_1' \text{ and} \\ C &\equiv B - (BD_2^b)[(BD_2^b)'(BD_2^b)]^-(BD_2^b)' \end{aligned}$$

the appropriate transformation for model (2.14) now is

$$P^b \equiv C - (CD_3)[(CD_3)'(CD_3)]^-(CD_3)' \tag{3.4}$$

It can easily be verified that P^b is idempotent and $P^bD^b = 0$, so all the fixed effects are indeed eliminated.

It is worth noting that both transformations (3.3) and (3.4) are dealing in a natural way with the no-self-flow problem, as only the rows corresponding to the $i = j$ observations need to be deleted from the corresponding dummy variables matrixes (in the unbalanced case, in fact from the D_1, D_2^a and D_1, D_2^b, D_3 matrixes [3]).

Transformation (3.4) can also be re-written in scalar form. First, let

$$\bar{\phi}^b \equiv \left(Q^b\right)^- \left(\bar{D}^b\right)' y \quad \text{where} \quad Q^b \equiv \left(D_2^b\right)' \bar{D}^b \quad \text{and} \quad \bar{D}^b \equiv \left(I_T - D_1(D_1'D_1)^{-1}D_1'\right) D_2^b,$$

$$\bar{\omega} \equiv \tilde{Q}^-(CD_3)'y \quad \text{where} \quad \tilde{Q} \equiv (CD_3)'(CD_3)$$

[3] Let use make a remark here. From a computational point of view the calculation of matrix B, more precisely $D_1(D_1'D_1)^{-1}D - 1'$ is by far the most resource requiring. Simplifications related to this can reduce dramatically CPU and Storage requirements.This topic, however, is beyond the limits of this paper, and the expertise of the authors.

and lastly

$$\bar{\xi} \equiv \left(Q^b\right)^- \left(\bar{D}^b\right)' D_3 \bar{\omega}$$

Now the scalar representation of transformation (3.4) is

$$\left[P^b y\right]_{ijt} = y_{ijt} - \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} y_{ijt} + \frac{1}{T_{ij}} \left(a_{ij}^b\right)' \bar{\phi}^b - \bar{\phi}_{it}^b - \bar{\omega}_{jt} + \frac{1}{T_{ij}} \tilde{a}_{ij}' \bar{\omega} + \bar{\xi}_{it} - \frac{1}{T_{ij}} \left(a_{ij}^b\right)' \bar{\xi}$$

where a_{ij}^b and $\tilde{a}_{ij}$ are the column vectors corresponding to observations ij from matrixes $A^b \equiv \left(D_2^b\right)' D_1$ and $\tilde{A} \equiv D_3' D_1$ respectively. $\bar{\phi}_{it}^b$ is the it-th element of the $(NT^* \times 1)$ column vector, $\bar{\phi}^b$. $\bar{\omega}_{jt}$ is the jt-th element of the $(NT^* \times 1)$ column vector, $\bar{\omega}$, and finally, $\bar{\xi}_{it}$ is the element corresponding to the it-th observation from the $(NT^* \times 1)$ column vector, $\bar{\xi}$.

Chapter 4

Dynamic Models

In the case of dynamic autoregressive models, the use of which is unavoidable if the data generating process has partial adjustment or some kind of memory, the Within estimators in a usual panel data framework are biased. In this section we generalize these well known results to this higher dimensional setup. We derive the finite sample bias for each of the models introduced in Chapter 2.

In order to show the problem, let us start with the simple linear dynamic model with bilateral interaction effects, that is model (2.5)

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \varepsilon_{ijt} \tag{4.1}$$

With backward substitution we get

$$y_{ijt} = \rho^t y_{ij0} + \frac{1-\rho^t}{1-\rho}\gamma_{ij} + \sum_{k=0}^{t} \rho^k \varepsilon_{ijt-k} \tag{4.2}$$

and

$$y_{ijt-1} = \rho^{t-1} y_{ij0} + \frac{1-\rho^{t-1}}{1-\rho}\gamma_{ij} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{ijt-1-k}$$

What needs to be checked is the correlation between the right hand side variables of model (4.1) after applying the appropriate Within transformation, that is the correlation between $(y_{ijt-1} - \bar{y}_{ij-1})$ where $\bar{y}_{ijt-1} = 1/T\sum_t y_{ijt-1}$ and $(\varepsilon_{ijt} - \bar{\varepsilon}_{ij})$ where $\bar{\varepsilon}_{ij} = 1/T\sum_t \varepsilon_{ij}$. This amounts to check the correlations $(y_{ijt-1}\bar{\varepsilon}_{ij})$, $(\bar{y}_{ij-1}\varepsilon_{ijt})$ and $(\bar{y}_{ij-1}\bar{\varepsilon}_{ij})$ because $(y_{ijt-1}\varepsilon_{ijt})$ are uncorrelated. These correlations are obviously not zero, not even in the semi-asymptotic case when $N \to \infty$, as we are facing the so called Nickell-type bias (*Nickell*[1981]). This may be the case for all other Within transformations as well.

Model (4.1) can of course be expanded to have exogenous explanatory variables as well

$$y_{ijt} = \rho y_{ijt-1} + x'_{ijt}\beta + \gamma_{ij} + \varepsilon_{ijt} \tag{4.3}$$

Let us turn now to the derivation of the finite sample bias and denote in general any of the above Within transformations by $\bar{y}_{trans}$. Using this notation we can derive the general form of the bias using *Nickell–type* calculations. Starting from the simple first order autoregressive model (4.1) introduced above we get

$$(y_{ijt} - \bar{y}_{trans}) = \rho(y_{ijt-1} - \bar{y}_{trans-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{trans}) \tag{4.4}$$

Using OLS to estimate ρ, we get

$$\widehat{\rho}_t = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1}) \cdot (y_{ijt} - \bar{y}_{trans})}{\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ijt-1} - \bar{y}_{trans-1})^2} \tag{4.5}$$

So in the expectations we have

$$E(\widehat{\rho} - \rho) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(y_{ijt-1} - \bar{y}_{trans-1}\right)\left(\varepsilon_{ijt} - \bar{\varepsilon}_{trans}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(y_{ijt-1} - \bar{y}_{trans-1}\right)^2} \tag{4.6}$$

Continuing with model (4.1) and using now the appropriate (2.6) Within transformation we get

$$(y_{ijt} - \bar{y}_{ij}) = \rho(y_{ijt-1} - \bar{y}_{ij-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij})$$

For the numerator of the bias in (4.6) from above we get

$$E[y_{ijt-1}\varepsilon_{ijt}] = 0$$

$$E[y_{ijt-1}\bar{\varepsilon}_{ij}] = E\left[\left(\sum_{k=0}^{t-1}\rho^k \varepsilon_{ijt-1-k}\right) \cdot \left(\frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{ijt}\right)\right] = \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho}$$

$$E[\bar{y}_{ij-1}\varepsilon_{ijt}] = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)\cdot(\varepsilon_{ijt})\right] = \frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{T-t}}{1-\rho}$$

$$E[\bar{y}_{ij-1}\bar{\varepsilon}_{ij}] = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)\cdot\left(\frac{1}{T}\cdot\sum_{t=1}^{T}\varepsilon_{ijt}\right)\right] = \frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)$$

And for the denominator of the bias in (4.6)

$$E[y_{ijt-1}^2] = E\left[\left(\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)^2\right] = \sigma_\varepsilon^2\cdot\frac{1-\rho^{2t}}{1-\rho^2}$$

$$\begin{aligned} E[y_{ijt-1}\bar{y}_{ij-1}] &= E\left[\left(\textstyle\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)\right] = \\ &= \tfrac{\sigma_\varepsilon^2}{T(1-\rho^2)}\left(\tfrac{1-\rho^t}{1-\rho}+\rho\tfrac{1-\rho^{T-t}}{1-\rho}-\rho^{t+1}\cdot\tfrac{1+\rho^T}{1-\rho}\right) \end{aligned}$$

$$\begin{aligned} E[\bar{y}_{ij-1}^2] &= E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\rho^k\varepsilon_{ijt-1-k}\right)^2\right] = \\ &= \tfrac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)}+\tfrac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right) \end{aligned}$$

So the finite sample bias for this model is

$$E\left[\hat{\rho}-\rho\right] = \frac{-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{t-1}}{1-\rho}\right)-\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1-\rho^{T-t}}{1-\rho}\right)+\frac{\sigma_\varepsilon^2}{T}\cdot\left(\frac{1}{1-\rho}-\frac{1}{T}\cdot\frac{1-\rho^T}{(1-\rho)^2}\right)}{\sigma_\varepsilon^2\cdot\left(\frac{1-\rho^{2t}}{1-\rho^2}\right)-A^*+B^*}$$

where

$$A^* = \frac{2\sigma_\varepsilon^2}{T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho}+\rho\frac{1-\rho^{T-t}}{1-\rho}-\rho^{t+1}\cdot\frac{1+\rho^T}{1-\rho}\right)$$

and

$$B^* = \frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\frac{2\rho(1-\rho^T)}{T(1-\rho^2)}+\frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)$$

It can be seen that these results are very similar to the original Nickell results, and the bias is persistent even in the semi-asymptotic case when $N \to \infty$.

Let us turn now our attention to model (2.1). In this case the Within transformation (2.3) leads to

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_t + \bar{y}) = \rho \cdot (y_{ijt-1} - \bar{y}_{ij-1} - \bar{y}_{t-1} + \bar{y}_{-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_t + \bar{\varepsilon})$$

After lengthy derivations (see the Appendix) we get for the finite sample bias

$$E\left[\hat{\rho} - \rho\right] = \frac{\left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{1-N^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\frac{1-\rho^{T-t}}{1-\rho} + A^*}{\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2} - B^* + C^*}$$

where

$$A^* = \left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T}\left(\frac{1}{1-\rho} - \frac{1}{T}\frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B^* = 2\left(\frac{N^2-1}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1}\cdot\frac{1+\rho^T}{1-\rho}\right)$$

and

$$C^* = \left(\frac{N^2-1}{N^2}\right)\frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

It is worth noticing that in the semi-asymptotic case as $N \to \infty$ we get back the bias derived above for model (4.1).

As seen earlier, the optimal Within transformation for model (2.1) is in fact (2.4)

$$(y_{ijt} - \bar{y}_i - \bar{y}_j - \bar{y}_t + 2\bar{y})$$

For this Within estimator the bias is (see the derivation in the Appendix)

$$E\left[\hat{\rho} - \rho\right] = \frac{\left(\frac{2-2N}{N^2}\right) \cdot \frac{\sigma_\epsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{2-2N}{N^2}\right) \cdot \frac{\sigma_\epsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho} + A^{**}}{\left(\frac{N^2-1}{N^2}\right) \cdot \sigma_\varepsilon^2 \frac{1-\rho^{2t}}{1-\rho^2} + B^{**} + C^{**}}$$

where

$$A^{**} = \left(\frac{2N-2}{N^2}\right) \cdot \frac{\sigma_\epsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B^{**} = \left(\frac{4-4N}{N^2}\right) \cdot \frac{\sigma_\epsilon^2}{T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right)$$

and

$$C^{**} = \left(\frac{2N-4}{N^2}\right) \frac{\sigma_\epsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

It can be seen as $N \to \infty$ the bias goes to zero, so this estimator is semi-asymptotically unbiased (unlike the previous one).

As the optimal Within transformation for model (2.7) is in fact (2.3), we get the same bias in this case as for model (2.1).

Let us now continue with models (2.8), (2.10) and (2.12) which can be considered as the same models from this point of view. Writing up model (2.8)

$$y_{ijt} = \rho y_{ijt-1} + \alpha_{jt} + \varepsilon_{ijt}$$

and applying the Within transformation to it we get

$$y_{ijt} - \bar{y}_{jt} = \rho \left(y_{ijt-1} - \bar{y}_{jt-1} \right) + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{jt} \right)$$

The bias then can be expressed as

$$E\left[\hat{\rho} - \rho\right] = \frac{E\left[y_{ijt-1}\varepsilon_{ijt}\right] - E\left[y_{ijt-1}\bar{\varepsilon}_{jt}\right] - E\left[\bar{y}_{jt-1}\varepsilon_{ijt}\right] + E\left[\bar{y}_{jt-1}\bar{\varepsilon}_{jt}\right]}{E\left[y_{ijt-1}^2\right] - 2 \cdot E\left[y_{ijt-1}\bar{y}_{jt-1}\right] + E\left[\bar{y}_{jt-1}^2\right]}$$

It can easily be seen that the expected value of the numerator is zero, as both y_{ijt-1} and $\bar{y}_{jt-1}$ depend on the ε-s only up to time $t-1$, and are necessarily uncorrelated with the t-th period disturbance, ε_{ijt}. So as the denominator is finite, the bias is in fact nil. The same arguments are valid for models (2.10) and (2.12) as well.

And finally, let us turn to model (2.14)

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt}$$

The Within transformation gives

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y})$$

so we get

$$\begin{aligned}(y_{ijt} - \bar{y}_{ij} & - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y}) = \\ & = \rho \cdot (y_{ijt-1} - \bar{y}_{ij-1} - \bar{y}_{jt-1} - \bar{y}_{it-1} + \bar{y}_{t-1} + \bar{y}_{j-1} + \bar{y}_{i-1} - \bar{y}_{-1}) + \\ & + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_{jt} - \bar{\varepsilon}_{it} + \bar{\varepsilon}_t + \bar{\varepsilon}_j + \bar{\varepsilon}_i - \bar{\varepsilon})\end{aligned}$$

And for the finite sample bias of this model we get

$$E[\hat{\rho} - \rho] = \frac{\left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho} + A^*}{\left(\frac{(N-1)^2}{N^2}\right) \sigma_\varepsilon^2 \frac{1-\rho^{2t}}{1-\rho^2} + B^* + C^*}$$

where

$$A^* = \left(\frac{(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T^2} \cdot \left(T \cdot \frac{1-\rho^{t-1}}{1-\rho} - \frac{\rho + (t-1)\rho^{t+1} - t\rho^t}{(1-\rho)^2}\right)$$

$$B^* = \left(\frac{-2(N-1)^2}{N^2}\right) \frac{\sigma_\varepsilon^2}{T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho \frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right)$$

and

$$C^* = \left(\frac{(N-1)^2}{N^2}\right) \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

It is clear that if N goes to infinity and T is finite, we get back the bias of model (2.5).

As seen above, we have problems with the estimation and N inconsistency of models (2.5), (2.7) and (2.14) in the dynamic case (see Table 2 in the appendix). Luckily, many of the well known instrumental variables (IV) estimators developed to deal with dynamic panel data models can be generalized to these higher dimensions as well, as the number of available orthogonality conditions increases together with the dimensions. Let us take the example of one of the most frequently used, the Arellano and Bond IV estimator (see *Arellano and Bond*[1991] and *Mark N. Harris and Sevestre*[2005] p. 260) for the estimation of model (2.5).

The model is written up in first differences, such as

$$y_{ijt} - y_{ijt-1} = \rho\left(y_{ijt-1} - y_{ijt-2}\right) + \left(\varepsilon_{ijt} - \varepsilon_{ijt-1}\right), \qquad t = 3, \ldots, T$$

or

$$\Delta y_{ijt} = \rho \Delta y_{ijt-1} + \Delta \varepsilon_{ijt}, \qquad t = 3, \ldots, T$$

The y_{ijt-k}, $(k = 2, \ldots, t-1)$ are valid instruments for Δy_{ijt-1}, as Δy_{ijt-1} is N asymptotically correlated with y_{ijt-k}, but y_{ijt-k} are not with $\Delta \varepsilon_{ijt}$. As a result, the full instrument set for a given cross sectional pair, (ij) is

$$z_{ij} = \begin{pmatrix} y_{ij1} & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ 0 & y_{ij1} & y_{ij2} & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & y_{ij1} & \cdots & y_{ijT-2} \end{pmatrix}_{((T-2)\times\frac{(T-1)(T-2)}{2})}$$

The resulting IV estimator of ρ is

$$\hat{\rho}_{AB} = \left[\Delta Y'_{-1} Z_{AB} \left(Z'_{AB}\Omega Z_{AB}\right)^{-1} Z'_{AB}\Delta Y_{-1}\right]^{-1} \Delta Y'_{-1} Z_{AB} \left(Z'_{AB}\Omega Z_{AB}\right)^{-1} Z'_{AB}\Delta Y,$$

where ΔY and ΔY_{-1} are the panel first differences, $Z_{AB} = [z'_{11}, z'_{12}, \ldots, z'_{NN}]'$ and $\Omega = I_{N^2} \otimes \Sigma$ is the covariance matrix, with known form

$$\Sigma = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}_{((T-2)\times(T-2))}$$

The generalized Arellano-Bond estimator is behaves exactly in the same way as the "original" two dimensional one, regardless the dimensionality of the model.

In the case of models (2.7) and (2.14) to derive an Arellano-Bond type estimator we need to insert one further step. After taking the first differences, we implement a simple transformation in order to get to a model with only (ij) pairwise interaction effects, exactly as in model (2.5). Then we proceed as above as the Z_{AB} instruments are going to be valid for these transformed models as well. Let us start with model (2.7) and take the first differences

$$y_{ijt} - y_{ijt-1} = \rho(y_{ijt-1} - y_{ijt-2}) + (\lambda_t - \lambda_{t-1}) + (\varepsilon_{ijt} - \varepsilon_{ijt-1})$$

Now, instead of estimating this equation directly with IV, we carry out the following transformation

$$(y_{ijt} \quad - y_{ijt-1}) - \frac{1}{N}\sum_{i=1}^{N}(y_{ijt} - y_{ijt-1}) = \rho\left[(y_{ijt-1} - y_{ijt-2}) - \frac{1}{N}\sum_{i=1}^{N}(y_{ijt-1} - y_{ijt-2})\right] + \\ + \left[(\lambda_t - \lambda_{t-1}) - \frac{1}{N}\sum_{i=1}^{N}(\lambda_t - \lambda_{t-1})\right] + \left[(\varepsilon_{ijt} - \varepsilon_{ijt-1}) - \frac{1}{N}\sum_{i=1}^{N}(\varepsilon_{ijt} - \varepsilon_{ijt-1})\right]$$

or introducing the notation $\Delta\tilde{y}_{jt} = \frac{1}{N}\sum_{i=1}^{N}(y_{ijt} - y_{ijt-1})$ and, also, noticing that the λ-s had been eliminated from the model

$$(\Delta y_{ijt} - \Delta\tilde{y}_{jt}) = \rho(\Delta y_{ijt-1} - \Delta\tilde{y}_{jt-1}) + (\Delta\varepsilon_{ijt} - \Delta\tilde{\varepsilon}_{jt})$$

We can see that the Z_{AB} instruments proposed above are valid again for $\Delta y_{ijt-1} - \Delta\tilde{y}_{jt-1}$ as well, as they are uncorrelated with $\Delta\varepsilon_{ijt} - \Delta\tilde{\varepsilon}_{jt}$, but correlated with the former. The

IV estimator of ρ, $\hat{\rho}_{AB}$ again has the form

$$\hat{\rho}_{AB} = \left[\Delta\tilde{Y}'_{-1} Z_{AB}(Z'_{AB}\Omega Z_{AB})^{-1} Z'_{AB}\Delta\tilde{Y}_{-1}\right]^{-1} \Delta\tilde{Y}'_{-1} Z_{AB}(Z'_{AB}\Omega Z_{AB})^{-1} Z'_{AB}\Delta\tilde{Y}.$$

Continuing now with model (2.14), the transformation needed in this case is

$$\begin{aligned}
\Delta y_{ijt} \quad & -\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta y_{ijt} - \tfrac{1}{N}\sum_{j=1}^{N}\Delta y_{ijt} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_{ijt} = \\
& = \rho\left[\Delta y_{ijt-1} - \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta y_{ijt-1} - \tfrac{1}{N}\sum_{j=1}^{N}\Delta y_{ijt-1} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_{ijt-1}\right] + \\
& + \left[\Delta\alpha_{it} - \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta\alpha_{it} - \tfrac{1}{N}\sum_{j=1}^{N}\Delta\alpha_{it} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta\alpha_{it}\right] + \\
& + \left[\Delta\alpha_{jt} - \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta\alpha_{jt} - \tfrac{1}{N}\sum_{j=1}^{N}\Delta\alpha_{jt} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta\alpha_{jt}\right] + \\
& + \left[\Delta\varepsilon_{ijt} - \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta\varepsilon_{ijt} - \tfrac{1}{N}\sum_{j=1}^{N}\Delta\varepsilon_{ijt} + \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta\varepsilon_{ijt}\right]
\end{aligned}$$

Picking up the previously introduced notation and using the fact that the fixed effects are cleared again we get

$$(\Delta y_{ijt} - \Delta\tilde{y}_{jt} - \Delta\tilde{y}_{it} + \Delta\tilde{y}_{t}) = \rho(\Delta y_{ijt-1} - \Delta\tilde{y}_{jt-1} - \Delta\tilde{y}_{it-1} + \Delta\tilde{y}_{t-1}) + (\Delta\varepsilon_{ijt} - \Delta\tilde{\varepsilon}_{jt} - \Delta\tilde{\varepsilon}_{it} + \Delta\tilde{\varepsilon}_{t})$$

The Z_{AB} instruments can be used again, on this transformed model, to get a consistent estimator for ρ.

Chapter 5

Further Extensions

We assumed so far throughout the paper that the idiosyncratic disturbance term ε is in fact a well behaved white noise, that is, all heterogeneity is introduced into the model through the fixed effects. In some applications this may be an unrealistic assumption, so next we relax it in two ways. We introduce heteroscedasticity and a simple form of cross correlation into the disturbance terms, and see how this impacts on the transformations introduced earlier. So far the approach has been to transform the models in such a way that the fixed effects drop out, and then estimate the transformed models with OLS. Now, however, after the appropriate transformation the model has to be estimated by Feasible GLS (FGLS) instead of OLS, as we have to take into account its covariance structure.

First, we derive the covariance matrix of the model and analyze how the different transformations introduced earlier impact on it. Then, we derive estimators for the variance components of the transformed model, in order to be able to use FGLS instead of

OLS for the estimation.

5.1 Covariance Matrixes and the Within Transformations

The initial Assumption (2.2) about the disturbance terms now is replaced by

$$E\left(\varepsilon_{ij}\varepsilon_{kl}\right)=\begin{cases}\sigma_{\varepsilon i}^{2} & \text{if } i=k, j=l, \forall t\\ \rho_{(1)} & \text{if } i=k, j\neq l, \forall t\\ \rho_{(2)} & \text{if } i\neq k, j=l, \forall t\\ 0 & \text{otherwise}\end{cases}$$

Then the covariance matrix of all models introduced in Section 2 takes the form

$$\Upsilon\equiv L_N\otimes I_{NT}-\left(\rho_{(1)}+\rho_{(2)}\right)\cdot I_{N^2T}+\rho_{(1)}\cdot I_N\otimes J_N\otimes I_T+\rho_{(2)}\cdot J_N\otimes I_N\otimes I_T,$$

where

$$L_N=\begin{pmatrix}\sigma_{\varepsilon 1}^{2} & 0 & \cdots & 0\\ 0 & \sigma_{\varepsilon 2}^{2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_{\varepsilon N}^{2}\end{pmatrix}_{N\times N}$$

This covariance matrix is altered, depending on the Within transformation used to get rid of the fixed effects.

In the case of transformation (2.3) the P_D projection matrix is

$$P_D = I_{N^2T} - \frac{1}{T} I_{N^2} \otimes J_T - \frac{1}{N^2} J_{N^2} \otimes I_T + \frac{1}{N^2T} J_{N^2T}$$

and we get

$$\begin{aligned} P_D \Upsilon P_D \quad &= \Upsilon - \tfrac{1}{T} L_N \otimes I_N \otimes J_T + \tfrac{1}{T}(\rho_{(1)} + \rho_{(2)}) \cdot I_{N^2} \otimes J_T \\ &-\tfrac{1}{N^2} L_N J_N \otimes J_N \otimes I_T - \tfrac{1}{N^2}\left((N-1)\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_{N^2} \otimes I_T \\ &+\tfrac{1}{N^2T} L_N J_N \otimes J_{NT} + \tfrac{1}{N^2T}\left((N-1)\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_{N^2T} \\ &-\tfrac{\rho_{(1)}}{T} \cdot I_N \otimes J_{NT} - \tfrac{\rho_{(2)}}{T} \cdot J_N \otimes I_N \otimes J_T \end{aligned}$$

For transformation (2.4) we get

$$P_D = I_{N^2T} - \frac{1}{NT} I_N \otimes J_{NT} - \frac{1}{NT} J_N \otimes I_N \otimes J_T - \frac{1}{N^2} J_{N^2} \otimes I_T + \frac{2}{N^2T} J_{N^2T}$$

and

$$
\begin{aligned}
P_D \Upsilon P_D &= \Upsilon - \tfrac{1}{NT} L_N \otimes J_{NT} - \tfrac{1}{NT}\left((N-1)\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_{NT} \\
&\quad - \tfrac{1}{NT} L_N J_N \otimes I_N \otimes J_T - \tfrac{1}{NT}\left(-\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_N \otimes I_N \otimes J_T \\
&\quad - \tfrac{1}{N^2} L_N J_N \otimes J_N \otimes I_T - \tfrac{1}{N^2}\left((N-1)\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_{N^2} \otimes I_T \\
&\quad + \tfrac{2}{N^2 T} L_N J_N \otimes J_T + \tfrac{1}{N^2 T}\left((N-2)\rho_{(1)} + (N-2)\rho_{(2)}\right) \cdot J_{N^2 T}
\end{aligned}
$$

For transformation (2.6) we have

$$P_D = I_{N^2 T} - \frac{1}{T} I_{N^2} \otimes J_T$$

and

$$
\begin{aligned}
P_D \Upsilon P_D &= \Upsilon - \tfrac{1}{T}\left(L_N \otimes I_N \otimes J_T\right) + \tfrac{1}{T}(\rho_{(1)} + \rho_{(2)}) \cdot I_{N^2} \otimes J_T - \tfrac{\rho_{(1)}}{T} \cdot I_N \otimes J_{NT} \\
&\quad - \tfrac{\rho_{(2)}}{T} \cdot J_N \otimes I_N \otimes J_T
\end{aligned}
$$

For transformation (2.9) we get

$$P_D = I_{N^2 T} - \frac{1}{N} J_N \otimes I_{NT}$$

and

$$
\begin{aligned}
P_D \Upsilon P_D &= \Upsilon - \tfrac{1}{N}\left(L_N J_N \otimes I_{NT}\right) + \tfrac{1}{N}(\rho_{(1)} + (1-N)\rho_{(2)}) \cdot J_N \otimes I_{NT} \\
&\quad - \tfrac{\rho_{(1)}}{N} \cdot J_{N^2} \otimes I_T
\end{aligned}
$$

For transformation (2.13) we get

$$P_D = I_{N^2T} - \frac{1}{N} J_N \otimes I_{NT} - \frac{1}{N} I_N \otimes J_N \otimes I_T + \frac{1}{N^2} J_{N^2} \otimes I_T$$

and

$$\begin{aligned} P_D \Upsilon P_D &= \Upsilon - \tfrac{1}{N} L_N J_N \otimes I_{NT} - \tfrac{1}{N}\left(-\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_N \otimes I_{NT} \\ &-\tfrac{1}{N} L_N \otimes J_N \otimes I_T - \tfrac{1}{N}\left((N-1)\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_N \otimes I_T \\ &+\tfrac{1}{N^2T} L_N J_N \otimes J_N \otimes I_T + \tfrac{1}{N^2}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_{N^2} \otimes I_T \end{aligned}$$

And finally, for transformation (2.15) we get

$$\begin{aligned} P_D &= I_{N^2T} - \tfrac{1}{N} J_N \otimes I_{NT} - \tfrac{1}{N} I_N \otimes J_N \otimes I_T - \tfrac{1}{T} I_{N^2} \otimes J_T \\ &+\tfrac{1}{NT} J_N \otimes I_N \otimes J_T + \tfrac{1}{NT} I_N \otimes J_{NT} + \tfrac{1}{N^2} J_{N^2} \otimes I_T - \tfrac{1}{N^2T} J_{N^2T} \end{aligned}$$

and

$$\begin{aligned} P_D \Upsilon P_D &= \Upsilon - \tfrac{1}{N} L_N J_N \otimes I_{NT} - \tfrac{1}{N}\left(-\rho_{(1)} + (N-1)\rho_{(2)}\right) \cdot J_N \otimes I_{NT} \\ &-\tfrac{1}{N} L_N \otimes J_N \otimes I_T - \tfrac{1}{N}\left((N-1)\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_N \otimes I_T \\ &+\tfrac{1}{N^2} L_N J_N \otimes J_N \otimes I_T + \tfrac{1}{N^2}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_{N^2} \otimes I_T \\ &-\tfrac{1}{T} L_N \otimes I_N \otimes J_T - \tfrac{1}{T}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot I_{N^2} \otimes J_T \\ &+\tfrac{1}{NT} L_N \otimes J_{NT} + \tfrac{1}{NT}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot I_N \otimes J_{NT} \\ &+\tfrac{1}{NT} L_N J_N \otimes I_N \otimes J_T + \tfrac{1}{NT}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_N \otimes I_N \otimes J_T \\ &-\tfrac{1}{N^2T} L_N J_N \otimes J_{NT} - \tfrac{1}{N^2T}\left(-\rho_{(1)} - \rho_{(2)}\right) \cdot J_{N^2T} \end{aligned}$$

5.2 Estimation of the Variance Components and the Cross Correlations

What now remains to be done is to estimate to variance components in order to make the GLS feasible. However, as we are going to see, so difficulties lay ahead. Let us start with the simples case, model (2.5). Applying transformation (2.6) leads to the following model to be estimated

$$(y_{ijt} - \bar{y}_{ij}) = (x_{ijt} - \bar{x}_{ij})\beta' + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij})$$

Let us denote the transformed disturbance terms by u_{ijt}. In this way now

$$E\left[u_{ijt}^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij})^2\right] = \frac{T-1}{T}\sigma_{\varepsilon i}^2$$

These are in fact N equations for N unknown parameters, so the system can be solved:

$$\begin{aligned} \tfrac{1}{NT} \quad & \textstyle\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 = \tfrac{T-1}{T}\hat{\sigma}_{\varepsilon i}^2 \\ \hat{\sigma}_{\varepsilon i}^2 \quad & = \tfrac{1}{N(T-1)}\textstyle\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 \end{aligned}$$

where $\hat{u}$ is the OLS residual from the estimation of the transformed model. We also have to estimate the two cross correlations, $\rho_{(1)}$ and $\rho_{(2)}$. This is done by taking the averages of the residuals with respect to i and j. Let us start with $\rho_{(1)}$

$$E\left[\left(\frac{1}{N}\sum_{j=1}^{N}u_{ijt}\right)^{2}\right]=E\left[\bar{u}_{it}^{2}\right]=\frac{T-1}{NT}\sigma_{\varepsilon i}^{2}+\frac{(N-1)(T-1)}{NT}\rho_{(1)}$$

As we already have an estimator for $\sigma_{\varepsilon i}^{2}$,

$$\begin{aligned}\hat{\rho}_{(1)} &= \tfrac{NT}{(N-1)(T-1)}\left(E\left[\hat{\bar{u}}_{it}^{2}\right]-\tfrac{T-1}{NT}\hat{\sigma}_{\varepsilon i}^{2}\right)= \\ &= \tfrac{1}{N^{2}(N-1)(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^{2}-\tfrac{1}{N(N-1)}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^{2}\end{aligned}$$

Now for $\rho_{(2)}$,

$$E\left[\left(\frac{1}{N}\sum_{i=1}^{N}u_{ijt}\right)^{2}\right]=E\left[\bar{u}_{jt}\right]=\frac{T-1}{N^{2}T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^{2}+\frac{(N-1)(T-1)}{NT}\rho_{(2)},$$

and so

$$\begin{aligned}\hat{\rho}_{(2)} &= \tfrac{NT}{(N-1)(T-1)}\left(E\left[\hat{\bar{u}}_{jt}^{2}\right]-\tfrac{T-1}{N^{2}T}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^{2}\right)= \\ &= \tfrac{1}{N^{2}(N-1)(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\hat{u}_{ijt}\right)^{2}-\tfrac{1}{N(N-1)}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^{2}\end{aligned}$$

For the other models the above exercise is slightly more complicated. Let us continue with model (2.1). For this model there were three transformations put forward in this paper, here we are using two of them (2.3) and (2.13):

$$\begin{aligned} E\left[u_{ijt}^2\right] &= E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_t + \bar{\varepsilon}\right)^2\right] = \\ &= \tfrac{(N^2-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \tfrac{T-1}{N^3T}\sum_{i=1}^N \sigma_{\varepsilon i}^2 - \tfrac{(N-1)(T-1)}{N^2T}\left(\rho_{(1)} + \rho_{(2)}\right) \\ E\left[u_{ijt}^{*2}\right] &= E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{it} - \bar{\varepsilon}_{jt} + \bar{\varepsilon}_t\right)^2\right] = \\ &= \tfrac{(N-2)(N-1)}{N^2}\sigma_{\varepsilon i}^2 + \tfrac{N-1}{N^3}\sum_{i=1}^N \sigma_{\varepsilon i}^2 - \tfrac{(N-1)^2}{N^2}\left(\rho_{(1)} + \rho_{(2)}\right) \end{aligned}$$

Let us notice notice that if we subtract $\frac{T-1}{T(N-1)}$ times the second equation from the first, we get

$$E\left[u_{ijt}^2\right] - \frac{T-1}{T(N-1)}E\left[u_{ijt}^{*2}\right] = -\frac{(N-1)^2}{N}\sigma_{\varepsilon i}^2$$

As a result

$$\hat{\sigma}_{\varepsilon i}^2 = -\frac{1}{(N-1)^2T}\sum_{j=1}^N\sum_{t=1}^T \hat{u}_{ijt}^2 + \frac{N(T-1)}{(N-1)^3T}\sum_{j=1}^N\sum_{t=1}^T \hat{u}_{ijt}^{*2}$$

Just as with the previous model, we can estimate $\rho_{(1)}$ and $\rho_{(2)}$ by taking the averages of the residuals. For $\rho_{(2)}$

$$\begin{aligned} E\left[\bar{u}_{jt}^2\right] &= E\left[\left(\bar{\varepsilon}_{jt} - \bar{\varepsilon}_j - \bar{\varepsilon}_t + \bar{\varepsilon}\right)^2\right] = \\ &= \tfrac{(N-1)(T-1)}{N^3T}\sum_{i=1}^N \sigma_{\varepsilon i}^2 - \tfrac{(N-1)(T-1)}{N^2T}\rho_{(1)} + \tfrac{(N-1)^2(T-1)}{N^2T}\rho_{(2)} \end{aligned}$$

Now we are ready to express $\hat{\rho}_{(2)}$

$$\frac{(N-1)(T-1)}{NT}\rho_{(2)} = \left[E\left[\bar{u}_{jt}^2\right] - \frac{T-1}{(N-1)T}\frac{1}{N}\sum_{i=1}^N E\left[u_{ijt}^{*2}\right]\right]$$

This leads to

$$\begin{aligned}\hat{\rho}_{(2)} &= \tfrac{NT}{(N-1)(T-1)}\left[\tfrac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\tfrac{1}{N}\sum_{i=1}^{N}\hat{u}_{ijt}\right)^2 - \tfrac{T-1}{N^2(N-1)T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}\right] = \\ &= \tfrac{1}{N^2(N-1)(T-1)}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\hat{u}_{ijt}\right)^2 - \tfrac{1}{N(N-1)^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}\end{aligned}$$

Doing the same for $\rho_{(1)}$ gives

$$\begin{aligned}E\left[\bar{u}_{it}^2\right] &= E\left[(\bar{\varepsilon}_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon})^2\right] = \tfrac{(N-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \tfrac{T-1}{N^3T}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \\ &\quad -\tfrac{(N-1)(T-1)}{N^2T}\rho_{(2)} + \tfrac{(N-1)^2(T-1)}{N^2T}\rho_{(1)}\end{aligned}$$

And so

$$\frac{(N-1)(T-1)}{NT}\rho_{(1)} = \left[E\left[\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{it}^2\right] - \frac{T-1}{(N-1)T}\frac{1}{N}\sum_{i=1}^{N}E\left[u_{ijt}^{*2}\right]\right]$$

which leads to

$$\begin{aligned}\hat{\rho}_{(1)} &= \tfrac{NT}{(N-1)(T-1)}\left[\tfrac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tfrac{1}{N}\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \tfrac{T-1}{N^2(N-1)T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}\right] = \\ &= \tfrac{1}{N^2(N-1)(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \tfrac{1}{N(N-1)^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^{*2}\end{aligned}$$

Let us continue with model (2.7). In this case we need to use two new Within transformations, and calculate the variances of the resulting transformed disturbance terms

(denoted by u^a and u^b)

$$\begin{aligned} E\left[(u^a_{ijt})^2\right] &= E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_{it} + \bar{\varepsilon}_i)^2\right] = \tfrac{(N-1)(T-1)}{NT}\sigma^2_{\varepsilon i} - \tfrac{(N-1)(T-1)}{NT}\rho_{(1)} \\ E\left[(u^b_{ijt})^2\right] &= E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_{jt} + \bar{\varepsilon}_j)^2\right] = \\ &= \tfrac{(N-2)(T-1)}{NT}\sigma^2_{\varepsilon i} + \tfrac{T-1}{N^2T}\textstyle\sum_{i=1}^N \sigma^2_{\varepsilon i} - \tfrac{(N-1)(T-1)}{NT}\rho_{(2)} \end{aligned}$$

Now, in order to express $\rho_{(1)}$ from the equations one has to transform further u^b_{ijt} by taking the averages with respect to j, and then take the average of the obtained variances with respect to i

$$\begin{aligned} \tfrac{1}{N}\textstyle\sum_{i=1}^N E\left[(\bar{u}^b_{it})^2\right] &= \tfrac{1}{N}\textstyle\sum_{i=1}^N E\left[(\bar{\varepsilon}_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon})^2\right] = \\ &= \tfrac{(N-1)(T-2)}{N^3T}\textstyle\sum_{i=1}^N \sigma^2_{\varepsilon i} + \tfrac{(N-1)^2(T-2)}{N^2T}\rho_{(1)} - \tfrac{(N-1)(T-2)}{N^2T}\rho_{(2)} \end{aligned}$$

It can be noticed that

$$\begin{aligned} \hat{\rho}_{(1)} &= \tfrac{N^2T}{(N-1)^2(T-2)}\left\{\tfrac{1}{N}\textstyle\sum_{i=1}^N E\left[(\bar{\hat{u}}^b_{it})^2\right] - \tfrac{(T-2)}{N^2(T-1)}\textstyle\sum_{i=1}^N E\left[(\hat{u}^b_{ijt})^2\right]\right\} = \\ &= \tfrac{1}{N(N-1)^2(T-2)}\textstyle\sum_{i=1}^N\sum_{t=1}^T\left(\sum_{j=1}^N \hat{u}^b_{ijt}\right)^2 - \tfrac{1}{N(N-1)^2(T-1)}\textstyle\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T(\hat{u}^b_{ijt})^2 \end{aligned}$$

The other components can easily be derived

$$\hat{\sigma}^2_{\varepsilon i} = \hat{\rho}_{(1)} + \frac{NT}{(N-1)(T-1)}E\left[(\hat{u}^a_{ijt})^2\right] = \hat{\rho}_{(1)} + \frac{1}{(N-1)(T-1)}\sum_{j=1}^N\sum_{t=1}^T(\hat{u}^a_{ijt})^2$$

$$\begin{aligned}\hat{\rho}_{(2)} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2 - \frac{NT}{(N-1)(T-1)}E\left[(\hat{u}_{ijt}^b)^2\right] = \\ &= \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2 - \frac{1}{N(N-1)(T-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}(\hat{u}_{ijt}^b)^2\end{aligned}$$

Unfortunately, we are not that lucky with the other models. The difficulty is that we are not able to transform the residuals or come forward with other transformations, which produce new, linearly independent equations to estimate the variance components. Instead we need to impose further restrictions on the models. Let us assume from now on that $\rho_{(1)} = \rho_{(2)} = \rho$.

For model (2.8) we have

$$E\left[u_{ijt}^2\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{jt})^2\right] = \frac{N-2}{N}\sigma_{\varepsilon i}^2 + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 - \frac{N-1}{N}\rho$$

Just like before, taking averages of u_{ijt} with respect to j leads to

$$E\left[\bar{u}_{it}^2\right] = E\left[(\bar{\varepsilon}_{it} - \bar{\varepsilon}_{t})^2\right] = \frac{N-2}{N^2}\sigma_{\varepsilon i}^2 + \frac{1}{N^3}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 + \frac{(N-1)(N-2)}{N^2}\rho$$

In this way we can estimate ρ

$$\begin{aligned}\hat{\rho} &= \frac{N^2}{(N-1)^2}\left[E\left[\bar{\hat{u}}_{it}^2\right] - \frac{1}{N}E\left[\hat{u}_{ijt}^2\right]\right] = \\ &= \frac{N}{(N-1)^2T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \frac{1}{N(N-1)^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2 = \\ &= \frac{1}{N(N-1)^2T}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\hat{u}_{ijt}\right)^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{u}_{ijt}^2\right\}\end{aligned}$$

and now we can move to estimate $\sigma^2_{\varepsilon i}$

$$\begin{aligned}
\hat{\sigma}^2_{\varepsilon i} &= \tfrac{N^2}{N-2}\left\{E\left[\bar{\hat{u}}^2_{it}\right] - \tfrac{1}{N^2(N-1)}\sum_{i=1}^N E\left[\hat{u}^2_{ijt}\right] - \tfrac{(N-1)(N-2)+1}{N^2}\hat{\rho}\right\} - = \\
&= \tfrac{N^2}{(N-2)T}\sum_{t=1}^T\left(\tfrac{1}{N}\sum_{j=1}^N \hat{u}_{ijt}\right)^2 - \tfrac{1}{N(N-1)(N-2)T}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{u}^2_{ijt} - \\
&\qquad - \tfrac{(N-1)(N-2)+1}{N-2}\hat{\rho} = \\
&= \tfrac{1}{(N-2)T}\sum_{t=1}^T\left(\sum_{j=1}^N \hat{u}_{ijt}\right)^2 - \tfrac{1}{N(N-1)(N-2)T}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{u}^2_{ijt} - \\
&\qquad - \tfrac{(N-1)(N-2)+1}{N-2}\hat{\rho}
\end{aligned}$$

Let us continue next with model (2.10). Now we have

$$E\left[u^2_{ijt}\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{it})^2\right] = \frac{N-1}{N}\sigma^2_{\varepsilon i} - \frac{N-1}{N}\rho$$

We can transform u_{ijt} further by taking the averages with respect to i and then compute the respective variances

$$E\left[\bar{u}^2_{jt}\right] = E\left[(\bar{\varepsilon}_{jt} - \bar{\varepsilon}_t)^2\right] = \frac{N-1}{N^3}\sum_{i=1}^N \sigma^2_{\varepsilon i} + \frac{(N-1)(N-2)}{N^2}\rho$$

So we can estimate ρ as

$$\begin{aligned}
\hat{\rho} &= \tfrac{N^2}{(N-1)^2}\left[E\left[\bar{\hat{u}}^2_{jt}\right] - \tfrac{1}{N^2}\sum_{i=1}^N E\left[\hat{u}^2_{ijt}\right]\right] = \\
&= \tfrac{N}{(N-1)^2 T}\sum_{j=1}^N\sum_{t=1}^T\left(\tfrac{1}{N}\sum_{i=1}^N \hat{u}_{ijt}\right)^2 - \tfrac{1}{N(N-1)^2 T}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{u}^2_{ijt} = \\
&= \tfrac{1}{N(N-1)^2 T}\left\{\sum_{j=1}^N\sum_{t=1}^T\left(\sum_{i=1}^N \hat{u}_{ijt}\right)^2 - \sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{u}^2_{ijt}\right\}
\end{aligned}$$

and $\sigma^2_{\varepsilon i}$ as

$$\hat{\sigma}^2_{\varepsilon i} = \hat{\rho} + \frac{1}{(N-1)T} \sum_{j=1}^{N} \sum_{t=1}^{T} \hat{u}^2_{ijt}$$

We still have two models, namely (2.12) and (2.14), to deal with. For these, however, unfortunately it is not possible to estimate any cross correlation at all. So we have to assume zero cross correlation and focus only on the heteroscedasticity and the estimation of the $\sigma^2_{\varepsilon i}$ variances.

For model (2.12) we have

$$E\left[u^2_{ijt}\right] = E\left[(\varepsilon_{ijt} - \bar{\varepsilon}_{it} - \bar{\varepsilon}_{jt} + \bar{\varepsilon}_t)^2\right] = \frac{(N-1)(N-2)}{N^2}\sigma^2_{\varepsilon i} + \frac{N-1}{N^3}\sum_{i=1}^{N} \sigma^2_{\varepsilon i}$$

Taking the averages with respect to i

$$\frac{1}{N}\sum_{i=1}^{N} E\left[u^2_{ijt}\right] = \frac{(N-1)^2}{N^3}\sum_{i=1}^{N} \sigma^2_{\varepsilon i}$$

As a result,

$$\begin{aligned} \hat{\sigma}^2_{\varepsilon i} &= \tfrac{N^2}{(N-1)^2(N-2)} \left\{ (N-1)E\left[\hat{u}^2_{ijt}\right] - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} E\left[\hat{u}^2_{ijt}\right] \right\} = \\ &= \tfrac{N}{(N-1)(N-2)T} \textstyle\sum_{j=1}^{N}\sum_{t=1}^{T} \hat{u}^2_{ijt} - \tfrac{1}{(N-1)^2(N-2)T} \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} \hat{u}^2_{ijt} \end{aligned}$$

We can proceed likewise for model (2.14)

$$
\begin{aligned}
E\left[u_{ijt}^2\right] &= E\left[\left(\varepsilon_{ijt} - \bar{\varepsilon}_{it} - \bar{\varepsilon}_{jt} - \bar{\varepsilon}_{ij} + \bar{\varepsilon}_i + \bar{\varepsilon}_j + \bar{\varepsilon}_t - \bar{\varepsilon}\right)^2\right] = \\
&= \tfrac{(N-1)(N-2)(T-1)}{N^2T}\sigma_{\varepsilon i}^2 + \tfrac{(N-1)(T-1)}{N^3T}\textstyle\sum_{i=1}^N \sigma_{\varepsilon i}^2
\end{aligned}
$$

Again,

$$
\frac{1}{N}\sum_{i=1}^N E\left[u_{ijt}^2\right] = \frac{(N-1)^2(T-1)}{N^3T}\sum_{i=1}^N \sigma_{\varepsilon i}^2
$$

and as a result,

$$
\begin{aligned}
\hat{\sigma}_{\varepsilon i}^2 &= \tfrac{N^2T}{(N-1)^2(N-2)(T-1)}\left\{(N-1)E\left[\hat{u}_{ijt}^2\right] - \tfrac{1}{N}\textstyle\sum_{i=1}^N E\left[\hat{u}_{ijt}^2\right]\right\} = \\
&= \tfrac{N}{(N-1)(N-2)(T-1)}\textstyle\sum_{j=1}^N\sum_{t=1}^T \hat{u}_{ijt}^2 - \tfrac{1}{(N-1)^2(N-2)(T-1)}\textstyle\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{u}_{ijt}^2
\end{aligned}
$$

Let us stop here a minute to make one last comment in this chapter. In previous chapters it is shown that many data problems may influence the consistency of the Within estimators. It would be nice to carry out the GLS (FGLS) estimator for such models if the no self-flow or the unbalanced data problem is present. Unfortunately we have no such luck estimating the variance components. Two things can happen: there is not enough identifying condition to express the variance components, or the underlying equations are so complex that it can not be computed by hand. However there is good news as well: Models (2.5), (2.8) and (2.10) are unaffected by both data problems, consequently the estimation of the variance components shall be done exactly as before.

Chapter 6

The 4-dimensional Setup

So far ijt-type observations had been treated, denoting for example an export activity from country i to country j at time t. In international trade however, it is more and more common to meet firm- or product level data. This extra information encourages researchers to modify the traditional 3-way ijt setup to a more suitable $ijst$-type observations. Naturally, this more detailed setup can be used to investigate economic problems in depths or even to address new ones.

This chapter is designed to introduce the most frequently used 4-dimensional fixed effects panel models and their Within transformations. Just as before, two common data problems are also treated, namely the lack of self trade and unbalanced data. Two conclusions will be drawn. Firstly, the fixed effects are generally depending on just a few indexes, meaning that usually multiple Within estimators can be chosen. Secondly, for the same reason, it is very well possible that the models can be transformed in such a way that the bias due to the data problem is eliminated.

6.1 The Models and the Within Transformations

The empirical meaning of such $ijst$-type observations is for example export from country i to country j of firm s (or of product s) at time t. Both i and j are assumed to go from $1 \ldots N$, as they denote the same set of countries involved in trade; s goes from 1 to some N_s and finally, t goes from $1 \ldots T$. Let us now have a closer look at the most relevant models and their optimal Within transformations. *Davis*[2002]'s method can be used to obtain such optimal Within transformations, even though in 4-dimensions this involves several "iterating" steps and the lengthy manipulation of more complex matrix forms. This procedure is illustrated for model (6.9), included in the appendix. Just as noted before, the optimal Within estimator is actually coincides with the LSDV estimator but as the number of parameters to be estimated grows rapidly, LSDV becomes unfeasible. It is more convenient then to transform the model first and then estimate it with OLS, giving the Within estimator.

The model extension of the traditional $\alpha_i + \lambda_t$ fixed effects structure takes the form

$$y_{ijst} = \beta' x_{ijst} + \alpha_{ijs} + \lambda_t + \varepsilon_{ijst}, \tag{6.1}$$

where y and x stand for the dependent- and the set of explanatory variables respectively and ε is the usual vector of disturbances. We can remove them with a simple linear transformation

$$y_{ijst} - \bar{y}_t - \bar{y}_{ijs} + \bar{y}, \tag{6.2}$$

where

$$\begin{aligned}
\bar{y}_t &= \tfrac{1}{N^2 N_s} \textstyle\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{N_s} y_{ijst} \\
\bar{y}_{ijs} &= \tfrac{1}{T} \textstyle\sum_{t=1}^{T} y_{ijst} \\
\bar{y} &= \tfrac{1}{N^2 N_s T} \textstyle\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst},
\end{aligned}$$

and naturally the x-s should be transformed in the same way. One can also come up with several other transformations as well, but (6.2) is the suggested due its optimality.

The next model

$$y_{ijst} = \beta' x_{ijst} + \alpha_i + \alpha_j + \alpha_s + \lambda_t + \varepsilon_{ijst} \tag{6.3}$$

captures all indexes in a separate fixed effect. The optimal transformation which wipes them out is

$$y_{ijst} - \bar{y}_i - \bar{y}_j - \bar{y}_s - \bar{y}_t + 3\bar{y}, \tag{6.4}$$

where

$$\begin{aligned}
\bar{y}_i &= \tfrac{1}{N N_s T} \textstyle\sum_{j=1}^{N} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst} \\
\bar{y}_j &= \tfrac{1}{N N_s T} \textstyle\sum_{i=1}^{N} \sum_{s=1}^{N_s} \sum_{t=1}^{T} y_{ijst} \\
\bar{y}_s &= \tfrac{1}{N^2 T} \textstyle\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} y_{ijst}.
\end{aligned}$$

One possible extension of the 3-dimensional bilateral Model (2.5) is

$$y_{ijst} = \beta' x_{ijst} + \alpha_{ij} + \alpha_{is} + \alpha_{js} + \varepsilon_{ijst} \tag{6.5}$$

with its Within transformation

$$y_{ijst} - \bar{y}_{js} - \bar{y}_{is} - \bar{y}_{ij} + \bar{y}_s + \bar{y}_j + \bar{y}_i - \bar{y}, \tag{6.6}$$

where

$$\begin{aligned} \bar{y}_{js} &= \tfrac{1}{NT} \textstyle\sum_{i=1}^{N} \sum_{t=1}^{T} y_{ijst} \\ \bar{y}_{is} &= \tfrac{1}{NT} \textstyle\sum_{j=1}^{N} \sum_{t=1}^{T} y_{ijst} \\ \bar{y}_{ij} &= \tfrac{1}{N_s T} \textstyle\sum_{s=1}^{N} \sum_{t=1}^{T} y_{ijst}. \end{aligned}$$

Also, we can present a variant of the above model as

$$y_{ijst} = \beta' x_{ijst} + \alpha_{ij} + \alpha_{is} + \alpha_{js} + \lambda_t + \varepsilon_{ijst}, \tag{6.7}$$

and its optimal transformation

$$y_{ijst} - \bar{y}_{js} - \bar{y}_{is} - \bar{y}_{ij} + \bar{y}_s + \bar{y}_j + \bar{y}_i - \bar{y}_t. \tag{6.8}$$

One can also come up with the extension of Model (2.12) in the form of

$$y_{ijst} = \beta' x_{ijst} + \alpha_{it} + \alpha_{jt} + \alpha_{st} + \varepsilon_{ijst}. \tag{6.9}$$

The optimal transformation clearing out all the fixed effects parameters is

$$y_{ijst} - \bar{y}_{it} - \bar{y}_{jt} - \bar{y}_{st} + 2\bar{y}_{t}, \tag{6.10}$$

where

$$\begin{aligned}
\bar{y}_{it} &= \tfrac{1}{NN_s} \textstyle\sum_{j=1}^{N} \sum_{s=1}^{N_s} y_{ijst} \\
\bar{y}_{jt} &= \tfrac{1}{NN_s} \textstyle\sum_{i=1}^{N} \sum_{s=1}^{N_s} y_{ijst} \\
\bar{y}_{st} &= \tfrac{1}{N^2} \textstyle\sum_{i=1}^{N} \sum_{j=1}^{N} y_{ijst}.
\end{aligned}$$

The last but possibly the most important model extension is the all encompassing model

$$y_{ijst} = \beta' x_{ijst} + \alpha_{ijs} + \alpha_{it} + \alpha_{jt} + \alpha_{st} + \varepsilon_{ijst}. \tag{6.11}$$

The optimal (and only) Within transformation which successfully cancels out all the fixed effects is

$$y_{ijst} - \bar{y}_{ijs} - \bar{y}_{jst} - \bar{y}_{ist} - \bar{y}_{ijt} + \bar{y}_{st} + \bar{y}_{jt} + \bar{y}_{js} + \bar{y}_{it} + \bar{y}_{is} + \bar{y}_{ij} - \bar{y}_{t} - \bar{y}_{s} - \bar{y}_{j} - \bar{y}_{i} + \bar{y}, \tag{6.12}$$

where

$$\begin{aligned}\bar{y}_{jst} &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} y_{ijst}\\ \bar{y}_{ist} &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N} y_{ijst}\\ \bar{y}_{ijt} &= \tfrac{1}{N_s}\textstyle\sum_{s=1}^{N_s} y_{ijst}.\end{aligned}$$

Let us stop here for a minute to stress out once more why it is so important to propose the Within estimators for all the above models. Just as in the 3-dimensional space but even more significantly, the number of parameters to be estimated is huge and computationally unfeasible: in case of model (6.1) for example, we should estimate $k + N^2 N_s + T$ parameters with LSDV (k is the number of explanatory variables), but only k with the Within estimator. This number is incredibly high even for moderate individual sizes.

6.2 No Self-Flow

Just as in the 3-dimensional case, it may well happen that within-country trade flows are unobservable and so missing from the database. From our point of view this means that for each jst we lack observations where $i = j$. We have already seen that several models had lost the unbiased and consistent properties of their respective estimators, but it was also shown, that there exist possible transformations (though not optimal), with which the remaining fixed effects are successfully wiped out. As it turns out, working in the 4-dimensional space brings even better results: as the lack of self-flow affects through only i and j, an extra index s is now present which also can be used to "average out" the fixed effects (remember that in the 3-dimensional case only t operated as this "extra" index). For models (6.1), (6.3), (6.11) the optimal Within transformations can be applied and so the Within estimators produce unbiased and consistent estimators. The problematic models are (6.5), (6.7), (6.9) (to see this why, check chapter 3), but coming up with other transformations treats the problem. For model (6.5) the transformation which addresses the no self-flow problem is as simple as

$$y_{ijst} - \bar{y}_{ijs}. \tag{6.13}$$

For model (6.7) transformation (6.2) can be applied and finally for model (6.9) coming up with transformation

$$y_{ijst} - \bar{y}_{jst} - \bar{y}_{ijt} + \bar{y}_{jt} \tag{6.14}$$

solves the problem and it is easy to see that all the remaining fixed effects are in fact dropped out. Going further up with the dimensions, the lack of self trade causes less and less trouble as the additional indexes can be used to average out the fixed effects regardless of the incomplete nature of the data.

6.3 Unbalanced Data

Just as with traditional panel data, it is uncommon to find complete datasets in the real world, where all individuals are observed through the exact same time scope. It would be a more realistic setup if all ijs observations are measured for $1 \ldots T_{ijs}$, rather than some general T. Just for this section it is assumed that $\sum_i \sum_j \sum_s T_{ijs} \equiv T$. It is important to note that $T_{ijs} = 0$ for many ijs individual (it is very unlikely for example that each ij country pair trades with every s product), however this does not affect the upcoming results. In the previous chapters it has already been shown that the unbalanced nature of the data left untouched some of the models, while for others either other transformation had to be presented or where it was impossible, the results of *Wansbeek and Kapteyn*[1989] had to be generalized. Our findings are somewhat similar to the no self-flow case in a sense that moving to higher dimensions improves the results: most models and their respective transformations are either perfectly usable in the unbalanced case or it is rather simple to propose a proper transformation. Let us see them in details. For model (6.9) unbalanced data does not cause any problem. For model (6.5) just like in the no self-flow case transformation (6.13) should be used whereas for models (6.3) and (6.7) the same transformation can be applied taking the form

$$y_{ijst} - \bar{y}_{jst} - \bar{y}_{ist} - \bar{y}_{ijt} + \bar{y}_{st} + \bar{y}_{jt} + \bar{y}_{it} - \bar{y}_t. \tag{6.15}$$

Unfortunately this is not the case for models (6.1) and (6.11). It is impossible to find such nice linear transformations, instead the method proposed by *Wansbeek and Kapteyn*[1989] should be applied. Let us now formalize it first for model (6.1), then for model (6.11). One should keep in mind that just as in their original paper, index t goes "slowly" and ijs fast (in their original order), allowing to model the *Wansbeek and Kapteyn*[1989] projection matrix. The dummy matrices $\tilde{D}_1^c = l_T^* \otimes I_{N^2 N_s}$ and $\tilde{D}_2^c = I_T^* \otimes l_{N^2 N_s}$ should be then modified in the following way to reflect the unbalanced nature of the data. Let R_t-s be $I_{N^2 N_s}$ matrices for each t, where we do the following: for all ijs observation, the rows corresponding to observation ijs are left untouched in the first T_{ijs} matrix but are deleted from the rest $T^* - T_{ijs}$ matrices, where $T^* \equiv max\{T_{ijs}\}$. In that way we define D_1^c and D_2^c as

$$\begin{aligned} D_1^c &\equiv [R_1', R_2', \dots, R_{T^*}']' \qquad (T \times N^2 N_s) \\ D_2^c &\equiv diag\{R_1, R_2, \dots, R_{T^*}\} \qquad (T \times T^*). \end{aligned}$$

Furthermore we introduce the following notations

$$\Delta_{N^2 N_s} \equiv D_1^{c'} D_1^c, \qquad \Delta_{T^*}^c \equiv D_2^{c'} D_2^c, \qquad A^c \equiv D_2^{c'} D_1^c$$

and

$$\begin{aligned} \bar{D}^c &\equiv D_2^c - D_1^c \Delta_{N^2 N_s}^{-1} A^{c'} = \left(I_T - D_1^c (D_1^{c'} D_1^c)^{-1} D_1^{c'}\right) D_2^c \\ Q^c &\equiv \Delta_{T^*}^c - A^c \Delta_{N^2 N_s}^{-1} A^{c'} = D_2^{c'} \bar{D}^c. \end{aligned}$$

Now with these formulas in hand we are ready to define the projection matrix

$$P^c = (I_T - D_1^c \Delta_{N^2 N_s}^{-1} D_1^{c'}) - \bar{D}^c Q^{c-} \bar{D}^{c'}, \tag{6.16}$$

where Q^{c-} denotes the generalized inverse of Q^c, as it has no full rank in general. For computational convenience it is important to also present the scalar version of the above transformation.

$$y_{ijst} - \frac{1}{T_{ijs}} \sum_{t=1}^{T_{ijs}} y_{ijst} - \bar{\phi}_t^c + \frac{1}{T_{ijs}} a_{ijs}^{c'} \bar{\phi}^c,$$

where $\bar{\phi}^c \equiv Q^{c-} \bar{D}^{c'} y$ and a_{ijs}^c denotes the column representing the ijs observation of matrix A^c (note that A^c has $N^2 N_s$ columns). Notice the huge difference between the two formulas: with the matrix notation one has to compute the inverse of an $N^2 N_s \times N^2 N_s$ matrix which might be computationally costly or even unfeasible, but picking up the scalar notation we only have to calculate a $T^* \times T^*$ inverse which is easily manageable.

Now let us turn our attention to the other problematic model, model (6.11). Essentially the same procedure has to be done, only it will be slightly more compelling due to the number of dummy matrices. Following the above notation, the dummy matrices $\tilde{D}_1^c = l_T^* \otimes I_{N^2 N_s}$, $\tilde{D}_2^d = I_T^* \otimes I_N \otimes l_{NN_s}$, $\tilde{D}_3^d = I_T^* \otimes l_N \otimes I_N \otimes l_{N_s}$ and $\tilde{D}_4^d = I_T^* \otimes l_{N^2} \otimes I_{N_s}$ should be modified to address the unbalanced nature of the data. Let us define $V_t^d = I_N \otimes l_{NN_s}$, $W_t^d = l_N \otimes I_N \otimes l_{N_s}$ and $U_t^d = l_N \otimes I_N \otimes l_{N_s}$ with the following modifications: for all 3 set of matrices, for all ijs, the rows corresponding to observation ijs should be left untouched in the first T_{ijs} matrix but be removed from the rest $T^* - T_{ijs}$. With this in hand, the new dummy matrices can be defined which will be used to set up the projection matrix later on:

$$\begin{aligned}
D_1^c &\equiv [R_1', R_2', \ldots, R_{T^*}']' \qquad (T \times N^2 N_s) \\
D_2^d &\equiv diag\left\{V_1^d, V_2^d, \ldots, V_{T^*}^d\right\} \qquad (T \times NT^*) \\
D_3^d &\equiv diag\left\{W_1^d, W_2^d, \ldots, W_{T^*}^d\right\} \qquad (T \times NT^*) \\
D_4^d &\equiv diag\left\{U_1^d, U_2^d, \ldots, U_{T^*}^d\right\} \qquad (T \times N_s T^*)
\end{aligned}$$

The projection matrix P^d should be built up in 4 steps using partial projection matrices in the following way:

$$\begin{aligned}
B^a &\equiv I_T - D_1^c (D_1^{c'} D_1^c)^{-1} D_1^{c'} \\
B^b &\equiv B^a - (B^a D_2^d)[(B^a D_2^d)'(B^a D_2^d)]^- (B^a D_2^d)' \\
B^c &\equiv B^b - (B^b D_3^d)[(B^b D_3^d)'(B^b D_3^d)]^- (B^b D_3^d)'
\end{aligned}$$

and finally the appropriate transformation is the outcome of the 4th step in the form

$$P^d = B^c - (B^c D_4^d)[(B^c D_4^d)'(B^c D_4^d)]^- (B^c D_4^d)'. \tag{6.17}$$

Summing up the results: just as in the 3-dimensional space, a very limited number of models are hard to be estimated consistently, but there exist tools readily available (though needed to be modified) to consistently estimate the problematic models as well. It should be noticed that in both 3- and 4-dimensions the same class of models perform badly; the coexistence of the all-individual fixed effect and some time-dependent one making the optimal Within estimator unbiased when the data is unbalanced. A summary containing the above results for all the models and their respective transformations can be found in Table 3 in the appendix.

Conclusion

This thesis was designed to deal with the most significant 3-way specifications of the fixed effects panel data models. It was easy to see that direct estimation of the models with LSDV could be very costly or even unfeasible as it involves the estimation of too many parameters. The obvious choice was then to use the Within estimator to deal with such fixed effects models. As it was shown, this generalization from the two-dimensional setup is not that straightforward and may involve rather complex algebra. It was also indicated that certain data problems can emerge while going up with the dimensions. Two had been analyzed in detail: the no self-flow, where the within-country observations are missing, and the unbalanced data case, where each individual observation is observed for a different time period. Some basic algebra showed that if these data related problems are present, several models can not be estimated consistently with the optimal Within estimator. To solve this, the following ideas had been presented. Firstly, as oppose to the two-way fixed effects models, where the Within transformations are mostly uniquely determined, in the three-way setup and further ahead there are multiple choices to operate with. Using an other Within transformation instead might avoid the data problem itself. Secondly, after some more complex calculus, more complex linear transformations could be found for the models where no self-flow still caused consistency problems. And finally, the application of *Wansbeek and Kapteyn*[1989]'s incomplete data results completed the analysis and left us with nothing less, but with all the models perfectly working from the consistency point of view. In later chapters dynamic models were examined and were concluded that the so-called *Nickell*[1981] bias is present in many models. To solve this, *Arellano and Bond*[1991] instrumental variable approach had been used, which led to the consistent estimation of all the corresponding models.

It was also obvious that the proposed models are quite flexible to further generalizations (four-dimensional fixed effect specifications had been just as easily analyzed as their three-dimensional counterparts) as well as to changes in the correlation structure of the disturbances.

It can be said then, that the thesis has successfully met its original goal: to give a guide to both theoretical and empirical researchers in how to work with the three- and four-way specifications of fixed effects panel data models. The main contribution of the thesis is that all the corresponding models can be estimated consistently even if some sort of data problem is present. As several different models had been taken into account, the thesis gives a heads up in how to estimate a new (not present here) model consistently. Higher dimensional generalizations are also possible (as it is shown in the thesis from three to four dimensions), hopefully opening up the way toward many nice future applications.

Appendix

Finite sample bias derivations for the dynamic model.

Model (2.1)

Now for model (2.1) transformation (2.3) leads to

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_t + \bar{y}) = \rho \cdot (y_{ijt-1} - \bar{y}_{ij-1} - \bar{y}_{t-1} + \bar{y}_{-1}) + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_t + \bar{\varepsilon})$$

Deriving the expected values, in the numerator we find

$$\begin{aligned}
E[y_{ijt-1}\varepsilon_{ijt}] &= E[y_{ijt-1}\bar{\varepsilon}_t] = E[\bar{y}_{t-1}\varepsilon_{ijt}] = E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0 \\
E[y_{ijt-1}\bar{\varepsilon}_{ij}] &= \tfrac{\sigma_\varepsilon^2}{T}\tfrac{1-\rho^{t-1}}{1-\rho} \\
E[y_{ijt-1}\bar{\varepsilon}] &= E[\bar{y}_{t-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{t-1}\bar{\varepsilon}] = \tfrac{\sigma_\varepsilon^2}{N^2T}\tfrac{1-\rho^{t-1}}{1-\rho} \\
E[\bar{y}_{ij-1}\varepsilon_{ijt}] &= \tfrac{\sigma_\varepsilon^2}{T}\tfrac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{ij-1}\bar{\varepsilon}_t] &= E[\bar{y}_{-1}\varepsilon_{ijt}] = E[\bar{y}_{-1}\bar{\varepsilon}_t] = \tfrac{\sigma_\varepsilon^2}{N^2T}\tfrac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{ij-1}\bar{\varepsilon}_{ij}] &= \tfrac{\sigma_\varepsilon^2}{T} \cdot \left(\tfrac{1}{1-\rho} - \tfrac{1}{T} \cdot \tfrac{1-\rho^T}{(1-\rho)^2}\right) \\
E[\bar{y}_{ij-1}\bar{\varepsilon}] &= E[\bar{y}_{-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{-1}\bar{\varepsilon}] = \tfrac{\sigma_\varepsilon^2}{N^2T} \cdot \left(\tfrac{1}{1-\rho} - \tfrac{1}{T} \cdot \tfrac{1-\rho^T}{(1-\rho)^2}\right)
\end{aligned}$$

and in the denominator

$$
\begin{aligned}
E[y^2_{ijt-1}] &= \sigma^2_\varepsilon \cdot \tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{t-1}] &= E[\bar{y}^2_{t-1}] = \tfrac{\sigma^2_\varepsilon}{N^2} \cdot \tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{ij-1}] &= \tfrac{\sigma^2_\varepsilon}{T(1-\rho^2)} \left(\tfrac{1-\rho^t}{1-\rho} + \rho\tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \tfrac{1+\rho^T}{1-\rho} \right) \\
E[y_{ijt-1}\bar{y}_{-1}] &= E[\bar{y}_{ij-1}\bar{y}_{t-1}] = \\
& E[\bar{y}_{t-1}\bar{y}_{-1}] = \tfrac{\sigma^2_\varepsilon}{N^2T(1-\rho^2)} \left(\tfrac{1-\rho^t}{1-\rho} + \rho\tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \tfrac{1+\rho^T}{1-\rho} \right) \\
E[\bar{y}^2_{ij-1}] &= \tfrac{\sigma^2_\varepsilon}{T(1-\rho)^2} \left(1 - \tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \tfrac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2} \right) \\
E[\bar{y}_{ij-1}\bar{y}_{-1}] &= E[\bar{y}^2_{-1}] = \tfrac{\sigma^2_\varepsilon}{N^2T(1-\rho)^2} \left(1 - \tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \tfrac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2} \right)
\end{aligned}
$$

The bias of this Within estimator for (2.1) is therefore the following

$$
E\left[\hat{\rho} - \rho\right] = \frac{\left(\frac{1-N^2}{N^2}\right)\frac{\sigma^2_\varepsilon}{T}\frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{1-N^2}{N^2}\right)\frac{\sigma^2_\varepsilon}{T}\frac{1-\rho^{T-t}}{1-\rho} + \left(\frac{N^2-1}{N^2}\right)\frac{\sigma^2_\varepsilon}{T^2} \cdot A^*}{\left(\frac{N^2-1}{N^2}\right) \cdot \sigma^2_\varepsilon \frac{1-\rho^{2t}}{1-\rho^2} - B^* + C^*}
$$

where

$$
A^* = \left(\frac{N^2-1}{N^2}\right)\frac{\sigma^2_\varepsilon}{T}\left(\frac{1}{1-\rho} - \frac{1}{T}\frac{1-\rho^T}{(1-\rho)^2}\right)
$$

$$
B^* = 2\left(\frac{N^2-1}{N^2}\right) \cdot \frac{\sigma^2_\varepsilon}{T(1-\rho^2)}\left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right)
$$

and

$$
C^* = \left(\frac{N^2-1}{N^2}\right)\frac{\sigma^2_\varepsilon}{T(1-\rho)^2}\left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)
$$

Now for the same model (2.1) transformation (2.4) leads to the following terms. For the numerator

$$
y_{ijt} - \bar{y}_i - \bar{y}_j - \bar{y}_t + 2\bar{y} = \rho\left(y_{ijt-1} - \bar{y}_{i-1} - \bar{y}_{j-1} - \bar{y}_{t-1} + 2\bar{y}_{-1}\right) + \left(\varepsilon_{ijt} - \bar{\varepsilon}_i - \bar{\varepsilon}_j - \bar{\varepsilon}_t + 2\bar{\varepsilon}\right)
$$

which yields the following terms. For the numerator

$$
\begin{aligned}
E[y_{ijt-1}\varepsilon_{ijt}] &= E[y_{ijt-1}\bar{\varepsilon}_t] = E[\bar{y}_{t-1}\varepsilon_{ijt}] = E[\bar{y}_{t-1}\bar{\varepsilon}_t] = 0 \\
E[y_{ijt-1}\bar{\varepsilon}_i] &= E[y_{ijt-1}\bar{\varepsilon}_j] = \tfrac{\sigma_\varepsilon^2}{NT}\tfrac{1-\rho^{t-1}}{1-\rho} \\
E[y_{ijt-1}\bar{\varepsilon}] &= E[\bar{y}_{t-1}\bar{\varepsilon}_i] = E[\bar{y}_{t-1}\bar{\varepsilon}_j] = E[\bar{y}_{t-1}\bar{\varepsilon}] = \tfrac{\sigma_\varepsilon^2}{N^2T}\tfrac{1-\rho^{t-1}}{1-\rho} \\
E[\bar{y}_{i-1}\varepsilon_{ijt}] &= E[\bar{y}_{j-1}\varepsilon_{ijt}] = \tfrac{\sigma_\varepsilon^2}{NT}\cdot\tfrac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{i-1}\bar{\varepsilon}_t] &= E[\bar{y}_{j-1}\bar{\varepsilon}_t] = E[\bar{y}_{-1}\varepsilon_{ijt}] = E[\bar{y}_{-1}\bar{\varepsilon}_t] = \tfrac{\sigma_\varepsilon^2}{N^2T}\cdot\tfrac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{i-1}\bar{\varepsilon}_i] &= E[\bar{y}_{j-1}\bar{\varepsilon}_j] = \tfrac{\sigma_\varepsilon^2}{NT}\cdot\left(\tfrac{1}{1-\rho} - \tfrac{1}{T}\cdot\tfrac{1-\rho^T}{(1-\rho)^2}\right) \\
E[\bar{y}_{i-1}\bar{\varepsilon}_j] &= E[\bar{y}_{j-1}\bar{\varepsilon}_i] = E[\bar{y}_{i-1}\bar{\varepsilon}] = E[\bar{y}_{j-1}\bar{\varepsilon}] = E[\bar{y}_{-1}\bar{\varepsilon}_i] = E[\bar{y}_{-1}\bar{\varepsilon}_j] = \\
&\quad E[\bar{y}_{-1}\bar{\varepsilon}] = \tfrac{\sigma_\varepsilon^2}{N^2T}\cdot\left(\tfrac{1}{1-\rho} - \tfrac{1}{T}\cdot\tfrac{1-\rho^T}{(1-\rho)^2}\right)
\end{aligned}
$$

and for the denominator

$$
\begin{aligned}
E[y_{ijt-1}^2] &= \sigma_\varepsilon^2\cdot\tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{t-1}] &= E[\bar{y}_{t-1}^2] = \tfrac{\sigma_\varepsilon^2}{N^2}\cdot\tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{i-1}] &= E[y_{ijt-1}\bar{y}_{j-1}] = \tfrac{\sigma_\varepsilon^2}{NT(1-\rho^2)}\left(\tfrac{1-\rho^t}{1-\rho} + \rho\tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1}\cdot\tfrac{1+\rho^T}{1-\rho}\right) \\
E[y_{ijt-1}\bar{y}_{-1}] &= E[\bar{y}_{i-1}\bar{y}_{t-1}] = E[\bar{y}_{j-1}\bar{y}_{t-1}] = \\
&\quad E[\bar{y}_{t-1}\bar{y}_{-1}] = \tfrac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)}\left(\tfrac{1-\rho^t}{1-\rho} + \rho\tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1}\cdot\tfrac{1+\rho^T}{1-\rho}\right) \\
E[\bar{y}_{i-1}^2] &= E[\bar{y}_{j-1}^2] = \tfrac{\sigma_\varepsilon^2}{NT(1-\rho)^2}\left(1 - \tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \tfrac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right) \\
E[\bar{y}_{i-1}\bar{y}_{-1}] &= E[\bar{y}_{j-1}\bar{y}_{-1}] = E[\bar{y}_{-1}^2] = \\
&\quad \tfrac{\sigma_\varepsilon^2}{N^2T(1-\rho)^2}\left(1 - \tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \tfrac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)
\end{aligned}
$$

Taking into account the sign and the frequency of the above elements the bias of this Within estimator is

$$
E\left[\hat{\rho}-\rho\right] = \frac{\left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{2-2N}{N^2}\right)\cdot\frac{\sigma_\varepsilon^2}{T}\cdot\frac{1-\rho^{T-t}}{1-\rho} + A^{**}}{\left(\frac{N^2-1}{N^2}\right)\cdot\sigma_\varepsilon^2\frac{1-\rho^{2t}}{1-\rho^2} + B^{**} + C^{**}}
$$

where

$$A^{**} = \left(\frac{2N-2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2}\right)$$

$$B^{**} = \left(\frac{4-4N}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho}\right)$$

and

$$\left(\frac{2N-4}{N^2}\right) \frac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \frac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \frac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2}\right)$$

Model (2.14)

Finally, let us turn to model (2.14)

$$y_{ijt} = \rho y_{ijt-1} + \gamma_{ij} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt}$$

The Within transformation gives

$$(y_{ijt} - \bar{y}_{ij} - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y}),$$

so we get

$$\begin{aligned}(y_{ijt} - \bar{y}_{ij} & - \bar{y}_{jt} - \bar{y}_{it} + \bar{y}_t + \bar{y}_j + \bar{y}_i - \bar{y}) = \\ & = \rho \cdot (y_{ijt-1} - \bar{y}_{ij-1} - \bar{y}_{jt-1} - \bar{y}_{it-1} + \bar{y}_{t-1} + \bar{y}_{j-1} + \bar{y}_{i-1} - \bar{y}_{-1}) + \\ & + (\varepsilon_{ijt} - \bar{\varepsilon}_{ij} - \bar{\varepsilon}_{jt} - \bar{\varepsilon}_{it} + \bar{\varepsilon}_t + \bar{\varepsilon}_j + \bar{\varepsilon}_i - \bar{\varepsilon})\end{aligned}$$

The expected value of the components in the numerator are the following

$$
\begin{aligned}
E[y_{ijt-1}\varepsilon_{ijt}] &= E[y_{ijt-1}\bar{\varepsilon}_{it}] = E[y_{ijt-1}\bar{\varepsilon}_{jt}] = E[y_{ijt-1}\bar{\varepsilon}_{t}] = E[\bar{y}_{it-1}\varepsilon_{ijt}] = E[\bar{y}_{jt-1}\varepsilon_{ijt}] = \\
E[\bar{y}_{it-1}\bar{\varepsilon}_{it}] &= E[\bar{y}_{jt-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{it-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{it-1}\bar{\varepsilon}_{t}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{t}] = \\
E[\bar{y}_{t-1}\varepsilon_{ijt}] &= E[\bar{y}_{t-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{t}] = 0 \\
E[y_{ijt-1}\bar{\varepsilon}_{ij}] &= \tfrac{\sigma_\varepsilon^2}{T} \cdot \tfrac{1-\rho^{t-1}}{1-\rho} \\
E[y_{ijt-1}\bar{\varepsilon}_{i}] &= E[y_{ijt-1}\bar{\varepsilon}_{j}] = E[\bar{y}_{it-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{it-1}\bar{\varepsilon}_{i}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{j}] = \\
&\tfrac{\sigma_\varepsilon^2}{NT} \cdot \tfrac{1-\rho^{t-1}}{1-\rho} \\
E[y_{ijt-1}\bar{\varepsilon}] &= E[\bar{y}_{it-1}\bar{\varepsilon}_{j}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{i}] = E[\bar{y}_{it-1}\bar{\varepsilon}_{j}] = E[\bar{y}_{jt-1}\bar{\varepsilon}_{i}] = E[\bar{y}_{it-1}\bar{\varepsilon}] = \\
E[\bar{y}_{jt-1} & \bar{\varepsilon}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{i}] = E[\bar{y}_{t-1}\bar{\varepsilon}_{j}] = E[\bar{y}_{t-1}\bar{\varepsilon}] = \tfrac{\sigma_\varepsilon^2}{N^2T} \cdot \tfrac{1-\rho^{t-1}}{1-\rho} \\
E[\bar{y}_{ij-1}\varepsilon_{ijt}] &= \tfrac{\sigma_\varepsilon^2}{T} \cdot \tfrac{1-\rho^{T-t}}{1-\rho}
\end{aligned}
$$

$$
\begin{aligned}
E[\bar{y}_{ij-1}\bar{\varepsilon}_{jt}] &= E[\bar{y}_{ij-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{i-1}\varepsilon_{ijt}] = E[\bar{y}_{j-1}\varepsilon_{ijt}] = E[\bar{y}_{i-1}\bar{\varepsilon}_{it}] = \\
E[\bar{y}_{j-1}\bar{\varepsilon}_{jt}] &= \tfrac{\sigma_\varepsilon^2}{NT} \cdot \tfrac{1-\rho^{T-t}}{1-\rho}
\end{aligned}
$$

$$
\begin{aligned}
E[\bar{y}_{ij-1}\bar{\varepsilon}_{t}] &= E[\bar{y}_{i-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{j-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{i-1}\bar{\varepsilon}_{t}] = E[\bar{y}_{j-1}\bar{\varepsilon}_{t}] = \\
E[\bar{y}_{-1}\varepsilon_{ijt}] &= E[\bar{y}_{-1}\bar{\varepsilon}_{jt}] = E[\bar{y}_{-1}\bar{\varepsilon}_{it}] = E[\bar{y}_{-1}\bar{\varepsilon}_{t}] = \tfrac{\sigma_\varepsilon^2}{N^2T} \cdot \tfrac{1-\rho^{T-t}}{1-\rho} \\
E[\bar{y}_{ij-1}\bar{\varepsilon}_{ij}] &= \tfrac{\sigma_\varepsilon^2}{T} \cdot \left(\tfrac{1}{1-\rho} - \tfrac{1}{T} \cdot \tfrac{1-\rho^T}{(1-\rho)^2} \right) \\
E[\bar{y}_{ij-1}\bar{\varepsilon}_{j}] &= E[\bar{y}_{ij-1}\bar{\varepsilon}_{i}] = E[\bar{y}_{i-1}\bar{\varepsilon}_{ij}] = E[\bar{y}_{j-1}\bar{\varepsilon}_{ij}] = \\
E[\bar{y}_{i-1}\bar{\varepsilon}_{i}] &= E[\bar{y}_{j-1}\bar{\varepsilon}_{j}] = \tfrac{\sigma_\varepsilon^2}{NT} \cdot \left(\tfrac{1}{1-\rho} - \tfrac{1}{T} \cdot \tfrac{1-\rho^T}{(1-\rho)^2} \right) \\
E[\bar{y}_{ij-1}\bar{\varepsilon}] &= E[\bar{y}_{i-1}\bar{\varepsilon}_{j}] = E[\bar{y}_{j-1}\bar{\varepsilon}_{i}] = E[\bar{y}_{i-1}\bar{\varepsilon}] = E[\bar{y}_{j-1}\bar{\varepsilon}] = E[\bar{y}_{-1}\bar{\varepsilon}_{ij}] = \\
E[\bar{y}_{-1}\bar{\varepsilon}_{i}] &= E[\bar{y}_{-1}\bar{\varepsilon}_{j}] = E[\bar{y}_{-1}\bar{\varepsilon}] = \tfrac{\sigma_\varepsilon^2}{N^2T} \cdot \left(\tfrac{1}{1-\rho} - \tfrac{1}{T} \cdot \tfrac{1-\rho^T}{(1-\rho)^2} \right)
\end{aligned}
$$

And in the denominator

$$
\begin{aligned}
E[y_{ijt-1}^2] &= \sigma_\varepsilon^2 \cdot \tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{it-1}] &= E[y_{ijt-1}\bar{y}_{jt-1}] = E[\bar{y}_{it-1}^2] = E[\bar{y}_{jt-1}^2] = \tfrac{\sigma_\varepsilon^2}{N} \cdot \tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{t-1}] &= E[\bar{y}_{it-1}\bar{y}_{jt-1}] = E[\bar{y}_{it-1}\bar{y}_{t-1}] = E[\bar{y}_{jt-1}\bar{y}_{t-1}] = E[\bar{y}_{t-1}^2] = \\
&\tfrac{\sigma_\varepsilon^2}{N^2} \cdot \tfrac{1-\rho^{2t}}{1-\rho^2} \\
E[y_{ijt-1}\bar{y}_{ij-1}] &= \tfrac{\sigma_\varepsilon^2}{T(1-\rho^2)} \left(\tfrac{1-\rho^t}{1-\rho} + \rho \tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \tfrac{1+\rho^T}{1-\rho} \right) \\
E[y_{ijt-1}\bar{y}_{i-1}] &= E[y_{ijt-1}\bar{y}_{j-1}] = E[\bar{y}_{ij-1}\bar{y}_{it-1}] = E[\bar{y}_{ij-1}\bar{y}_{jt-1}] = \\
E[\bar{y}_{it-1}\bar{y}_{i-1}] &= E[\bar{y}_{jt-1}\bar{y}_{j-1}] = \tfrac{\sigma_\varepsilon^2}{NT(1-\rho^2)} \left(\tfrac{1-\rho^t}{1-\rho} + \rho \tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \tfrac{1+\rho^T}{1-\rho} \right) \\
E[y_{ijt-1}\bar{y}_{-1}] &= E[\bar{y}_{ij-1}\bar{y}_{t-1}] = E[\bar{y}_{it-1}\bar{y}_{j-1}] = E[\bar{y}_{jt-1}\bar{y}_{i-1}] = E[\bar{y}_{it-1}\bar{y}_{-1}] = \\
E[\bar{y}_{jt-1}\bar{y}_{-1}] &= E[\bar{y}_{t-1}\bar{y}_{i-1}] = E[\bar{y}_{t-1}\bar{y}_{j-1}] = E[\bar{y}_{t-1}\bar{y}_{-1}] = \\
&\tfrac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)} \left(\tfrac{1-\rho^t}{1-\rho} + \rho \tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \tfrac{1+\rho^T}{1-\rho} \right) \\
E[\bar{y}_{ij-1}^2] &= \tfrac{\sigma_\varepsilon^2}{T(1-\rho)^2} \left(1 - \tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \tfrac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2} \right) \\
E[\bar{y}_{ij-1}\bar{y}_{i-1}] &= E[\bar{y}_{ij-1}\bar{y}_{j-1}] = E[\bar{y}_{i-1}^2] = E[\bar{y}_{j-1}^2] = \\
&\tfrac{\sigma_\varepsilon^2}{NT(1-\rho)^2} \left(1 - \tfrac{2\rho(1-\rho^T)}{T(1-\rho^2)} + \tfrac{2\rho^{T+2} - \rho^{2(T+1)} - \rho^2}{1-\rho^2} \right)
\end{aligned}
$$

$$
\begin{aligned}
E[\bar{y}_{ij-1}\bar{y}_{-1}] &= E[\bar{y}_{i-1}\bar{y}_{j-1}] = E[\bar{y}_{i-1}\bar{y}_{-1}] = E[\bar{y}_{j-1}\bar{y}_{-1}] = E[\bar{y}_{-1}^2] = \\
&\tfrac{\sigma_\varepsilon^2}{N^2T(1-\rho^2)} \left(\tfrac{1-\rho^t}{1-\rho} + \rho \tfrac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \tfrac{1+\rho^T}{1-\rho} \right)
\end{aligned}
$$

To sum up the bias we get for this model is

$$
E[\widehat{\rho} - \rho] = \frac{\left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{t-1}}{1-\rho} + \left(\frac{-(N-1)^2}{N^2}\right) \cdot \frac{\sigma_\varepsilon^2}{T} \cdot \frac{1-\rho^{T-t}}{1-\rho} + A^*}{\left(\frac{(N-1)^2}{N^2}\right) \sigma_\varepsilon^2 \frac{1-\rho^{2t}}{1-\rho^2} + B^* + C^*}
$$

where

$$
A^* = \frac{(N-1)^2}{N^2} \cdot \frac{\sigma_\varepsilon^2}{T} \left(\frac{1}{1-\rho} - \frac{1}{T} \cdot \frac{1-\rho^T}{(1-\rho)^2} \right)
$$

$$
B^* = \frac{-2(N-1)^2}{N^2} \cdot \frac{\sigma_\varepsilon^2}{T(1-\rho^2)} \left(\frac{1-\rho^t}{1-\rho} + \rho \frac{1-\rho^{T-t}}{1-\rho} - \rho^{t+1} \cdot \frac{1+\rho^T}{1-\rho} \right)
$$

and

$$C^* = \left(\frac{(N-1)^2}{N^2}\right)\frac{\sigma_\varepsilon^2}{T(1-\rho)^2}\left(1-\frac{2\rho(1-\rho^T)}{T(1-\rho^2)}+\frac{2\rho^{T+2}-\rho^{2(T+1)}-\rho^2}{1-\rho^2}\right)$$

Davis[2002]'s method for Model (6.9)

The dummy matrices are

$$\begin{aligned} D_1 = &\ I_N \otimes l_{NN_s} \otimes I_T \\ D_2 = &\ l_N \otimes I_N \otimes l_{N_s} \otimes I_T \\ D_3 = &\ l_{N^2} \otimes I_{N_sT}, \end{aligned}$$

where just as before I is identity matrix, l is the column of ones. According to the procedure the projection matrix can be obtained in the form of $P_D = Q_1 - P_2 - P_3$, where

$$\begin{aligned} Q_1 = &\ I_{N^2N_sT} - D_3(D_3'D_3)^{-1}D_3' \\ P_2 = &\ (Q_1D_2)[D_2'Q_1D_2]^{-1}(Q_1D_2)' \\ P_3 = &\ (Q_2Q_1D_1)[D_1'(Q_2Q_1)D_1]^{-1}(Q_2Q_1D_1)', \end{aligned}$$

where $Q_2 = I_{N^2N_sT} - P_2$. We can compute each of the three projection matrix in turn.

$$Q_1 = I_{N^2N_sT} - \bar{J}_{N^2} \otimes I_{N_sT},$$

where $\bar{J}$ stands as before.

$$P_2 = \bar{J}_N \otimes I_N \otimes \bar{J}_{N_s} \otimes I_T - \bar{J}_{N^2N_s} \otimes I_T,$$

and finally

$$P_3 = I_N \otimes \bar{J}_{NN_s} \otimes I_T - \bar{J}_{N^2N_s} \otimes I_T.$$

Putting together the results:

$$\begin{aligned}
P_D &= Q_1 - P_2 - P_3 \\
&= I_{N^2N_sT} - \bar{J}_{N^2} \otimes I_{N_sT} - \bar{J}_N \otimes I_N \otimes \bar{J}_{N_s} \otimes I_T + \bar{J}_{N^2N_s} \otimes I_T - \\
&- I_N \otimes \bar{J}_{NN_s} \otimes I_T + \bar{J}_{N^2N_s} \otimes I_T \\
&= I_{N^2N_sT} - \bar{J}_{N^2} \otimes I_{N_sT} - \bar{J}_N \otimes I_N \otimes \bar{J}_{N_s} \otimes I_T - I_N \otimes \bar{J}_{NN_s} \otimes I_T + 2\bar{J}_{N^2N_s} \otimes I_T,
\end{aligned}$$

equivalently in scalar form

$$y_{ijst} - \bar{y}_{st} - \bar{y}_{jt} - \bar{y}_{it} + 2\bar{y}_t.$$

Table 1: The Behavior of the Proposed Within Estimators for 3-way Models

Model		(2.1)			(2.5)	(2.7)		(2.8)	(2.10)	(2.12)		(2.14)		
Transformation		(2.4)	(2.3)	(2.13)	(2.6)	(2.3)	(3.3)	(2.9)	(2.11)	(2.13)	(3.1)	(2.15)	(3.2)	(3.4)
CD	**Finite** N, T	+	+	+	+	+	+	+	+	+	-	+	-	+
	$N \to \infty$	+	+	+	+	+	+	+	+	+	-	+	-	+
NSF	**Finite** N, T	-	+	-	+	+	+	+	+	-	+	-	+	+
	$N \to \infty$	+	+	+	+	+	+	+	+	+	+	+	+	+
UBD	**Finite** N, T	-	-	+	+	-	+	+	+	+	-	-	-	+
	$N \to \infty$	-	-	+	+	-	+	+	+	+	-	-	-	+

where + stands for no bias, *CD*, *NSF* and *UBD* stand for *Complete Data*, *No Self-Flow* and *Unbalanced Data* respectively.

Table 2: The Behavior of the Proposed Within Transformations in Case of Dynamic Models

Model	(2.1)		(2.5)	(2.7)	(2.8)	(2.10)	(2.12)	(2.14)
Transformation	(2.3)	(2.4)	(2.6)	(2.3)	(2.9)	(2.11)	(2.13)	(2.15)
Finite N, T	-	-	-	-	+	+	+	-
Finite T, $N \to \infty$	-	+	-	-	+	+	+	-
A-B GMM	+		+	+				+

where + stands for no bias, A-B for the *Arellano and Bond*[1991] GMM estimator

Table 3: The Behavior of the Proposed Within Estimators for 4-way Models

Model	(6.1)		(6.3)		(6.5)		(6.7)			(6.9)		(6.11)	
Transformation	(6.2)	(6.16)	(6.4)	(6.15)	(6.6)	(6.13)	(6.8)	(6.2)	(6.15)	(6.10)	(6.14)	(6.12)	(6.17)
CD (Finite)	+	+	+	+	+	+	+	+	+	+	+	+	+
NSF (Finite)	+	+	+	-	-	+	-	+	-	-	+	+	+
UBD (Finite)	-	+	-	+	-	+	-	-	+	+	-	-	+

where + stands for no bias, *CD*, *NSF* and *UBD* stand for *Complete Data*, *No Self-Flow* and *Unbalanced Data*.

Bibliography

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 127–134.

BAIER, S. L., AND J. H. BERGSTRAND (2007): "Do Free Trade Agreements Actually Increase Members International Trade?," *Journal of International Economics*, 71, 72–95.

BALDWIN, R., AND D. TAGLIONI (2006): "Gravity for Dummies and Dummies for the Gravity Equations," *NBER Working Paper 12516.*

BALESTRA, P., AND J. KRISHNAKUMAR (2008): *Fixed Effects and Fixed Coefficients Models, in Matyas and Sevestre (eds.); The Econometrics of Panel Data.* Sringer Verlag, 3rd edn.

BALTAGI, B. (1995): *Econometric Analysis of Panel Data.* John Wiley & Sons.

BALTAGI, B. H., P. EGGER, AND M. PFAFFERMAYR (2003): "A generalized Design for Bilateral Trade Flow Models," *Economic Letters*, 80, 391–397.

DAVIS, P. (2002): "Estimating multi-way error components models with unbalanced data structures," *Journal of Econometrics*, 106, 67–95.

Egger, P., and M. Pfaffermayr (2003): "The Proper Econometric Specification of the Gravity Equation: 3-Way Model with Bilateral Interaction Effects," *Empirical Economics*, 28, 571–580.

Gourieroux, C., and A. Monfort (1989): "Statistique et Modeles Econometriques," *Economica*, pp. 184–185.

Hornok, C. (2011): "Gravity or Dummies? The Limits of Identification in Gravity Estimations," *CeFig Working Papers*, 15.

Laszlo Matyas, M. N. H., and L. Konya (2011/05): "Within Transformations for Three-Way Fixed Effects Models of Trade," *unpublished manuscript.*

Mark N. Harris, L. M., and P. Sevestre (2005): *Dynamic Models for Short Panels; in Matyas and Sevestre (eds.) The Econometrics of Panel Data.* Springer Verlag, 3rd edn.

Matyas, L. (1997): "Proper Econometric Specification of the Gravity Model," *The World Economy*, 363–369, 20.

Matyas, L., and L. Balazsi (2011/12): "The Estimation of Multi-dimensional Fixed Effects Panel Data Models," *CEU Economics Working Papers.*

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

Wansbeek, T., and A. Kapteyn (1989): "Estimation of Error-Components Model With Incomplete Panels," *Journal of Econometrics*, 41, 341–361.